Array-based Autoantibody Profiling and Epitope Mapping

Arash Zandian

Doctoral Thesis

KTH Royal Institute of Technology
Kungliga Tekniska Högskolan
School of Biotechnology
Stockholm, Sweden 2017

KTH Royal Institute of Technology
Kungliga Tekniska Högskolan
School of Biotechnology
Division of Proteomics and Nanobiotechnology
Science for Life Laboratory
Tomtebodavägen 23A
171 65 Solna
Sweden

ISBN 978-91-7729-499-3
TRITA-BIO Report 2017:19
ISSN 1654-2312

Printed by US-AB 2017

Cover illustration: Sara Stjernlöf

"Science is the organized skepticism in the reliability of expert opinion"

Richard Feynman

Abstract

Antibodies are a class of proteins that are made by the immune system to recognize harmful organisms and molecules. Their exceptional capability of specifically recognizing molecules has been investigated for over a century and information thereof has been utilized for a variety of applications including vaccine and generation of therapeutic antibodies. Occasionally, instead of protecting the host against pathogens, antibodies can recognize constituents of the host and thereby cause an autoimmune reaction that eventually can lead to a disease. Therefore, it is of great interest to understand what the antibodies bind to and their specificities.

The last decades of technical development and availability of protein and peptide microarrays have enabled large-scale profiling of antibodies and precise determination of their specificities through epitope mapping. In this thesis the aim was to use affinity proteomics tools to profile antibodies, determine their specificities, and discover potential associations of autoantigens to disease by analyzing blood-derived samples with microarray-based methods.

In Paper I, 57 serum samples from patients with the suggested autoimmune disease narcolepsy, were analyzed on planar antigen microarrays with 10,846 human protein fragments. Verification on an independent sample collection consisting of serum samples from 176 individuals, revealed METTL22 and NT5C1A as two potential autoantigens. In **Paper II**, antibodies from 53 plasma samples from patients with first-episode psychosis, a condition suggested to have a partial autoimmune component, were analyzed on planar antigen microarrays with 2,304 human protein fragments. After a follow-up study of the patients, antibodies toward an antigen representing the three proteins, PAGE2, PAGE2B, PAGE5, was found associated to an increased risk of developing schizophrenia. In **Paper III**, serum and plasma samples from patients with the autoimmune diseases multiple sclerosis and narcolepsy, were epitope mapped on high-density peptide microarrays with approximately 2.2 million peptides. Technical and biological verification, by using other microarray technology and analyzing samples from 448 patients, revealed one peptide for multiple sclerosis and narcolepsy, representing the proteins MAP3K7 and NRXN1, with higher antibody reactivity towards in each group, respectively. In **Paper IV**, purified polyclonal antibodies raised against a surface antigen found on malaria-infected erythrocytes, were profiled on the

peptide microarrays representing all proteins found on malaria-infected erythrocytes derived from *Plasmodium falciparum*. Then, different Plasmodium falciparum strains were analyzed by immunofluorescence microscopy and western blots, using the epitope mapped antibodies. The performance of the immunoassays were compared to the identified epitopes, and validated by RNA sequencing.

In conclusion, these investigations describe multiplex methods to identify and characterize antibodies, their disease association and epitopes. Follow-up studies are needed to determine their potential use and clinical value.

Keywords: Antibody, antigens, peptide, epitope mapping, autoimmunity, autoantibodies, microarrays.

Populärvetenskaplig sammanfattning

Antikroppar är en grupp av protein som skapas av immunförsvaret, och har som uppgift att skydda oss mot skadliga molekyler och sjukdomsframkallande organismer som t.ex. virus och bakterier. Detta åstadkommer antikroppar genom att specifikt binda till ett så kallat antigen, som exempelvis kan vara ett protein på bakteriens cellvägg, som sedan markerar för kroppens immunförsvar att bryta ned den.

Antikroppar skapas kontinuerligt av immunförsvaret för att skydda oss mot alla framtida hot som kan ge upphov till skada, vilket gör att det finns miljontals olika antikroppar hos oss människor. Immunförsvaret kan också specifikt skapa antikroppar mot vissa antigen när kroppen utsätts för en infektion eller vid till exempel en vaccinering.

Men ibland kan immunförsvaret ta fel på vad som är kroppsfrämmande, vilket leder till att immunförsvaret bryter ner delar av den egna kroppen med hjälp av autoantikroppar (auto = själv), via en så kallad autoimmun reaktion. Detta kan så småningom leda till utvecklingen av en autoimmun sjukdom (exempelvis typ-1-diabetes). I **Paper I** och **Paper II** studerades narkolepsi respektive psykos, vilket är sjukdomar som har föreslagits vara autoimmuna. Med hjälp av blodplasma och serum från patienter med de sjukdomarna kan autoantikropparna analyseras och profileras mot kroppsegna proteiner. Detta gav en inblick kring vad det är för proteiner i kroppen som angrips av de eventuella autoantikropparna och slutligen identifierades ett fåtal proteiner som kan vara av betydelse för sjukdomarna.

Anledningen till varför autoimmuna sjukdomar bryter ut är omdebatterat och ännu inte helt klarlagt. En föreslagen mekanism är att det i vissa fall kan finnas en molekylär likhet mellan sjukdomsorsakande organismers proteiner, och kroppens egna proteiner, vilket exempelvis tros ha förorsakat narkolepsi vid massvaccineringen mot influensan H1N1 år 2009 i bland annat Sverige. Genom att ta reda på vad antikropparna i detalj binder till genom en så kallad epitopmappning, vilket görs i **Paper III,** inom narkolepsi och multipel skleros, kan detaljerad information fås om var på proteinerna de eventuella autoantikropparna binder. Detta resulterade i att två regioner på respektive

protein kunde identifieras som uppvisade skillnad i hur autoantikropparna band mellan friska och sjuka patienter.

Immunförsvarets förmåga att generera antikroppar mot skadliga molekyler har länge utnyttjats av forskare. På liknande vis som ett vaccin fungerar, kan man få djur att skapa antikroppar mot specifika antigen, exempelvis speciella proteiner man vill studera. I **Paper IV**, togs antikroppar fram på detta vis för att studera parasiten som orsakar malaria, *Plasmodium falciparum*. Antikropparna i **Paper IV** hade specifikt tagits fram mot ett visst protein som bara finns på ytan av röda blodkroppar när en viss stam av *Plasmodium falciparum* har infekterat dem. Infektion av andra stammar av *Plasmodium falciparum* ger upphov till andra typer av proteiner på cellytan. I studien studerades vad antikropparna i detalj band till med hjälp av epitopmappning för att relatera den informationen till antikropparnas bindningsförmågor till olika stammar av *Plasmodium falciparum* vid olika analysmetoder.

Thesis defense

This thesis will be defended October 6th 2017 at 10.00, in Air & Fire, Science For Life Laboratory, Tomtebodavägen 23A, Solna, for the degree of Doctor of Technology in Biotechnology (PhD in Biotechnology).

Respondent

Arash Zandian graduated as a Master of Science in Engineering from Kungliga Tekniska Högskolan (KTH) in 2012, and worked as a lab engineer before starting his PhD studies at the Division of Proteomics and Nanobiotechnology, KTH, in 2013.

Faculty Opponent

Morten Nielsen, Professor at Department of Bio and Health Informatics, Technical University of Denmark, Lyngby, & Professor at Instituto de Investigaciones Biotecnológicas, Universidad de San Martín, Buenos Aires, Argentina.

Evaluation Committee

Anna Rostedt Punga, Associate professor at Department of Neuroscience, Uppsala University.

Adnane Achour, Professor at Division for Infectious Diseases, Department of Medicine, Karolinska Institutet.

Johan Normark, Assistant Professor at Divison of Infectious Diseases, Department of Clinical Microbiology, Umeå University.

Chairman of the Thesis Defense

Mathias Uhlén, Professor at Division of Proteomics and Nanobiotechnology, Science for Life Laboratory, School of Biotechnology, KTH.

Main Supervisor

Peter Nilsson, Professor at Division of Proteomics and Nanobiotechnology, Science for Life Laboratory, School of Biotechnology, KTH.

Co-supervisors:

Jochen M Schwenk, Associate Professor at Division of Proteomics and Nanobiotechnology, Science for Life Laboratory, School of Biotechnology, KTH.

Maja Neiman, Researcher at Division of Proteomics and Nanobiotechnology, Science for Life Laboratory, School of Biotechnology, KTH.

List of publications

The presented thesis is based on the following four articles, referred to by their Roman numerals (I-IV).

Paper I

Autoantibody targets in vaccine-associated narcolepsy

Anna Häggmark-Månberg, **Arash Zandian**, Björn Forsström, Mohsen Khademi, Izaura Lima Bomfim, Cecilia Hellström, Lisen Arnheim-Dahlström, Tove Hallböök, Niklas Darin, Ingrid E. Lundberg, Mathias Uhlén, Markku Partinen, Jochen M. Schwenk, Tomas Olsson & Peter Nilsson.

Autoimmunity **2016,** 49, (6), 421-433. DOI: 10.1080/08916934.2016.1183655

Paper II

Untargeted screening for novel autoantibodies with prognostic value in first-episode psychosis

Arash Zandian*, Louise Wingård*, Hanna Nilsson, Evelina Sjöstedt, Daniel X Johansson, David Just, Cecilia Hellström, Mathias Uhlén, Jochen M Schwenk, Anna Häggmark-Månberg, Oscar Norbeck, Björn Owe-Larsson, Peter Nilsson & Mats AA Persson.

Transl Psychiatry **2017,** 7, (7), e1177. DOI: 10.1038/tp.2017.160

Paper III

Whole-proteome peptide microarrays for profiling autoantibody repertoires within multiple sclerosis and narcolepsy

Arash Zandian, Björn Forsström, Anna Häggmark-Månberg, Jochen M. Schwenk, Mathias Uhlén, Peter Nilsson & Burcu Ayoglu.

J Proteome Res **2017,** 16, (3), 1300-1314. DOI: 10.1021/acs.jproteome.6b00916

Paper IV

Epitopes of anti-RIFIN antibodies and characterization of *rif*-expressing *Plasmodium falciparum* parasites by RNA sequencing

Jun-Hong Ch´ng, Madle Sirel*, **Arash Zandian***, Maria del Pilar Quintana*, Sherwin Chan*, Kirsten Moll*, Asa Tellgren-Roth*, IngMarie Nilsson, Peter Nilsson, Ulrika Qundos, & Mats Wahlgren.

Sci Rep **2017,** 7, 43190. DOI: 10.1038/srep43190

** The authors contributed equally to the work*

Respondent's contributions to the included papers

Paper I

Partly responsible for experimental planning and experimental work, joint responsibility for data analysis and shared responsibility and contribution manuscript writing.

Paper II

Primarily responsible for experimental planning on the arrays, experimental work and subsequent image- and data analysis. Shared responsibility and contribution during manuscript writing.

Paper III

Jointly responsibility for experimental planning on the arrays and experimental work and primarily responsible for image- and data analysis. Shared responsibility and contribution during manuscript writing.

Paper IV

Jointly responsible for experimental planning and experimental work on the arrays, and primarily responsible for image and data analysis of the array data. Partially responsible for manuscript writing.

Related publications not included in the thesis

Protein Profiling in Plasma Reveals Molecular Subgroups in Systemic Lupus Erythematosus
Arash Zandian*, Helena Idborg*, Iva Gunnarsson, Elisabet Svenungsson, Peter Nilsson and Per-Johan Jakobsson
Manuscript in preparation (2017)

Molecular deconvolution of circulating fibulin-1 by affinity proteomics
Arash Zandian*, Sofia Bergström*, Claudia Fredolini, Mathias Uhlén, Jacob Odeberg, Jochen M. Schwenk, Peter Nilsson and Maja Neiman
Manuscript in preparation (2017)

Dysregulations in circulating sphingolipids associate with disease activity indices in female patients with systemic lupus erythematosus: a cross-sectional study
A Checa, H Idborg, **Arash Zandian**, D Garcia Sar, I Surowiec, J Trygg, E Svenungsson, P-J Jakobsson, P Nilsson, I Gunnarsson and CE Wheelock
Lupus **2017**, 961203316686707. DOI: 10.1177/0961203316686707

Antibodies to PfEMP1, RIFIN and SURFIN expressed at the *Plasmodium falciparum* parasitized red blood cell surface in children with malaria.
Maria del Pilar Quintana, Jun-Hong Ch´ng, **Arash Zandian,** Peter Nilsson, Somporn Saiwaew, Kirsten Moll, Ulrika Qundos and Mats Wahlgren
Submitted to Scientific Reports (2017)

Anoctamin 2 identified as an autoimmune target in multiple sclerosis
Burcu Ayoglu, Nicholas Mitsios, Ingrid Kockum, Mohsen Khademi, **Arash Zandian**, Ronald Sjöberg, Björn Forsström, Johan Bredenberg, Izaura Lima Bomfim, Erik Holmgren, Hans Grönlund, André Ortlieb Guerreiro-Cacais, Nada Abdelmagid, Mathias Uhlén, Tim Waterboer, Lars Alfredsson, Jan Mulder, Jochen M Schwenk, Tomas Olsson, Peter Nilsson
Proc Natl Acad Sci U S A **2016,** 113, (8). DOI: 10.1073/pnas.1518553113

**The authors contributed equally to the work*

Aims of the thesis

This thesis aims to identify and characterize antibodies, their disease association and epitopes, by using microarray-based methods. This was applied to antibodies obtained from patients with suggested autoimmune diseases, but also from purified polyclonal antibodies used for research.

Paper I - The aim of the study was to screen for potential autoantigens in patients with narcolepsy by profiling their IgG repertoire on protein fragment microarrays. The IgG reactivities were also compared in relation to etiology, i.e. sporadic or vaccine-associated nacolepsy.

Paper II - The aim of this study was to screen for potential autoantigens in patients first-episode psychosis. Furthermore, patients were followed for a mean duration of 7 years, to see if a subsequent diagnosis could be associated to molecular findings.

Paper III - The aim of this study was to epitope map the antibody repertoire, and reveal potential autoantigens and epitopes that can be associated to the autoimmune diseases multiple sclerosis.

Paper IV - The aim of this study was to epitope map polyclonal antibodies raised against a *Plasmodium falciparum* protein that is expressed on the surfaces of infected erythrocytes, to relate their epitopes to performances in different analyses.

Contents

Abstract i

Populärvetenskaplig sammanfattning iii

Thesis defense v

List of publications vi

Respondent's contributions to the included papers vii

Related publications not included in the thesis viii

Aims of the thesis ix

Contents xi

1. What is Life? 1

Understanding the building blocks of life 1

From DNA to proteins 6

2. Proteomics 9

Do we really need proteomics? 9

Protein structures 10

Studying the proteome 12

Affinity-based proteomics 13

Human Protein Atlas project 14

Mass-spectrometry-based methods 15

Challenges in studying the human proteome 16

3. Antibodies 20

Structure 20

Function 22

Achieving Diversity of Binding/Recognition 24

Antibodies – A toolbox for molecular biology 27

4. Antigens and Epitopes 32

Epitopes and Paratopes 32

Autoreactivity in autoimmune diseases 35

Epitope mapping and identification of autoantigens 40

Site-directed mutagenesis 41

Mass spectrometry-based methods 41

Array-based methods 42

Surface display-based methods 44

Prediction-based methods for epitope mapping antibodies 45

Present investigations 47
Paper I: Autoantibody targets in vaccine-associated narcolepsy 48
Paper II: Untargeted screening for novel autoantibodies with prognostic value in first-episode psychosis 50
Paper III: Whole-proteome peptide microarrays for profiling autoantibody repertoires within multiple sclerosis and narcolepsy 53
Paper IV: Epitopes of anti-RIFIN antibodies and characterization of *rif*-expressing *Plasmodium falciparum* parasites by RNA sequencing 56
Challenges and considerations 59
Concluding remarks 62
Acknowledgements 67
References 70

1. What is Life?

What is life? Ever since my high school biology teacher asked this question, I have been intrigued by its overwhelming complexity and depth. Although one might intuitively understand what life is, this deceivingly trivial question has currently no unanimous answer in the scientific community [1]. There have been many attempts in summarizing the key mechanisms of what life is, a recent definition suggests that life should: store information, improvise solutions, exist in a confined space, have a metabolism, regenerate its molecules, adapt to its environment, and seclude molecules [2].

In order to fully understand this definition and the underlying processes that make up life, we must understand its building blocks, i.e., the molecules of life.

In the following chapters of my thesis, I will first give a brief historical perspective to my work in the field of proteomics, and then specifically focus on a group of molecules produced by our immune system, the antibodies. I will review the role of antibodies, as they both protect our body against diseases by recognizing pathogenic molecules, as well as occasionally causing diseases. Finally, I will describe methods and applied work for characterizing what molecules these antibodies are interacting with.

Understanding the building blocks of life

In 1665, the microscopy pioneer Robert Hooke coined the term "cell". This referred to microscopic porous honeycomb structures that he observed in thinly sliced cork [3]. A decade later, Antonie van Leeuwenhoek had further developed the microscope, leading to the discoveries of bacteria, protists and red blood cells [4]. Although the smallest entity of life, the cell, was now discovered, scientists were most likely unaware of the significance of their work at the time.

Two centuries later, based on the findings of Hooke and Leeuwenhoek, significant key discoveries were made by Matthias Jacob Schleiden and Theodor Schwann [5]. In 1838 Schleiden showed that plants are built up of cells

and described them as the smallest entity of the plants. In 1839, Schwann showed that cells in animals were similar to the cells observed in plants and concluded that they must be the smallest entities of life. Furthermore, Schwann reported that cells can make up the diversity seen in the different tissues of a plant and animal, and concluded that each cell possesses an individual life.

Following the discovery of cells researchers started to further study their chemical makeup, though this had begun prior to Schleiden and Schwann published their cell theory.

In the late 18th century Antoine François de Fourcroy studied what was at that time referred to as animal substances. It arose from a belief that plants are formed from minerals, and animals from plants. In his book from 1794, Fourcroy describes that one of the main differences between animal substances and plant substances were their solid matters, which are made up from different textures of organs found in animals. He described that these differences are constituted by three different groups of substances: albumin, gelatin and fibrin, and that different combinations and proportions of these substances form all the different substances of animals [6].

In 1835, Gerardus Johannes Mulder started to investigate the animal substances that were previously described by Fourcroy, as well as some plant substances. Mulder found that the elemental compositions (carbon, hydrogen, nitrogen and oxygen) of the substance in wheat was almost identical to the animal substances albumin and fibrin [7]. Prior to publication of these results, Mulder summarized them in a letter and sent it to his colleague and friend Jöns Jacob Berzelius, who suggested that these unifying substances should be called protein (derived from the greek: *in the lead*) [8]. In Mulder's following publication in 1838 [9], he described these unifying substances as protein, and claimed: "without it no life appears possible on our planet".

The theory of evolution was the next important step towards today's view on the fundamentals of life. Charles Darwin presented this theory in the book "On the origin of species" in 1859 [10]. This book was the culmination of Darwin's observations during his voyage on the Beagle He described the origin of species as a process of natural selection, suggesting that all forms of life are of common descent. However, Darwin did not provide an answer to the mechanism of heredity [11] and it would be Gregor Mendel who published his work in 1865, where he studied pea plants in the context of artificial fertilization and

hybridization that would suggest the mechanisms of hereditary [12]. Contrary to the prevailing belief that traits are mixed and passed on to following generations, he showed evidence for heritance based on a regularity that could be predicted.

Unfortunately, Mendel's work did not get any significant attention by the research community until the early 20th century. In 1909 Wilhelm Johannsen coined the term "gene" (as well as the terms genotype and phenotype), which referred to the unknown elements that are heritable [13-15]. In the following years when the trio John Burdon Sanderson Haldane, Sewall Wright and Ronald Aylmer Fisher, independently contributed to combine Mendel's theories on hereditary with Darwin's view on natural selection and thus the unified theory of evolution was born [16,17], which still applies today, albeit with a few extensions [18].

Key findings were made by Friedrich Miescher, who in 1869 wrote about his attempts to understand the building blocks of the cell that determines its structure, i.e. the "molecules of life", by studying lymphoid cells obtained from pus. This led him to the discovery of a substance located in the nucleus of the cell that did not contain proteins, and in contrast to proteins, it had a reversible appearance when changing from acidic to alkaline conditions. This substance was subsequently named "nuclein" in 1871, due to its location in the cell [19]. Albrecht Kossel, Richard Altmann and Albert Neumann further characterized nuclein (later to be named nucleic acids), contributing to identification of its constituents [20]. Between the years 1885 and 1893, Kossel, determined that chromatin is a mix of proteins and nucleic acids [19,20]. In parallel to the molecular discoveries on chromatin, the cytologist Walter Sutton was studying the morphology of chromosomes in the spermatogonial divisions in grasshoppers. With contributions from the work of Theodor Boveri, in 1903 Sutton concluded the following: chromosomes differ qualitatively and represent distinct potentialities, they occur in two equivalent series (pairs), the chromosome pairs separates upon cell division, and chromosomes are the bases of hereditary characters [21,22].

The focus on nucleic acids as hereditary molecules was gradually reduced in the beginning of the 20th century. Scientists believed, due to the diversity seen in proteins and its constituents compared to the nucleic acids, proteins must be the carrier of hereditary information [20,23]. In addition, nucleic acid preparations were highly degraded, and thus their immense complexity was not

realized[23]. Furthermore, Phoebus Levene *et al.* had introduced the tetranucleotide hypothesis, a hypothesis in early 20th century that was not in the favor of the nucleic acids as hereditary molecules, since it stated that nucleic acids form tetramers and thus constraining its ability of carrying heritable information [20]. However, Levene's work had also set up a foundation for studying nucleic acids by having further characterized the nucleotides (guanine, adenine, cytosine, thymine and uracil), the sugar backbone (deoxyribose and ribose), and the phosphate group [24,25]. This eventually lead to the identification of ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) [26].

Interest in nucleic acids was rejuvenated in 1928 when Frederick Griffith published a paper on his study of strains of pneumococci to understand the virulence of pneumonia-causing bacteria By adding virulent pneumococci strains that had been killed by heat (100°C), to an avirulent strain and then into mice, Griffith could observe that the virulence and its characteristics could be transformed from the killed to the living strain [27].

An intense search to understand the transforming agent began, and by 1931; other research groups showed that it was possible to induce the transformation *in vitro*. The following year; another research group showed that the cell-free fluid from the heat-killed pneumococci could still induce transformation [28]. These findings paved the way for the work of Oswald Avery, Maclyn McCarty and Colin MacLeod that they published in 1944, where they methodically tested the conditions necessary for purifying the transforming agent. They were able to rule out the transforming agent as a protein, lipid or polysaccharide, showing that it must be a highly polymerized and viscous form of DNA [28].

In years following, much work was done to understand the nature of DNA. In 1950, Erwin Chargaff studied the molar composition of the nucleotides in nucleic acids from different animals and in bacteria. By observing the different molar ratios of the nucleotides, Chargaff supported the idea of DNA as responsible for transmission of inherited properties, and subsequently laid the first steps in understanding the base pairing of the nucleotides in DNA (Chargaff's rules) [29]. Finally, the work of Alfred Hershey and Martha Chase in 1952 on isotope-labeled bacteriophages showed that certain isotopes corresponding to DNA were incorporated in bacteria, while isotopes corresponding to proteins retained in the phage, and thus removed any doubt that DNA was the information carrier [30].

Finally, in 1953, the double helical structure of DNA was resolved using x-ray diffraction, fully supporting the hypothesis that DNA was a polymer and was the carrier of hereditary information [31]. This discovery is often accredited James Watson and Francis Crick, but many scientists contributed to the discovery, including Rosalind Franklin, Raymond Gosling, Maurice Wilkins, Linus Pauling and Robert Corey [20].

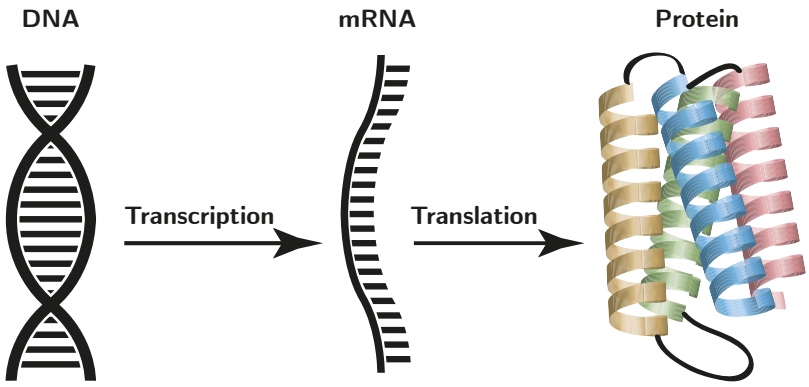


***Figure 1**. A simplistic view of the central dogma of molecular biology, DNA transcribes to mRNA that translates into protein.*

But how were DNA and the proteins that Mulder described in 1838 as "essential for life" connected? In 1958 Crick was the first to present what is known as the central dogma of biology, an explanation of how information, or as Crick defined it, "precise determination of sequences" flows between DNA to RNA to proteins [32], through processes today known as transcription and translation, Figure 1. Furthermore, a protein cannot pass information back to nucleic acids. Although Crick postulated that RNA is an intermediate molecule between DNA and protein, at that time there was no direct evidence to support this. In 1958 Lazarus Astrachan and Elliot Volkin performed isotope labeling with phosphor to track metabolic activity of RNA upon infecting *E. coli* with bacteriophages [33]. They discovered that if the isotopes are added shortly after phage infection, they will be incorporated primarily into RNA, but if the isotopes are added later, they will be incorporated into DNA. They failed, however, to describe a valid mechanism for their findings. Only three years later Sydney Brenner *et al.* [34] proposed a model for an unstable intermediate, derived from the genes that carries information about what protein to

synthesize. They followed the incorporation of isotopes in different compartments in the host upon bacteriophage infection over time, and could see that the new RNA that was formed had the same base composition as the phage DNA. Thus concluding that they had evidence for a messenger RNA (mRNA) that transports the genetic information from the nucleus to the ribosomes, to initiate protein synthesis.

From DNA to proteins

In the same article that the central dogma of biology was introduced in 1958, Crick discusses the importance of proteins. He first quoted his colleague Francis Watson: "The most significant thing about the nucleic acids is that we don't know what they do" and then contrasted this statement, saying: "the most significant thing about proteins is that they can do almost anything" [32]. These quotes both illustrates that research on proteins had progressed further than research on nucleic acids.

To put Crick's words into context; the first enzyme was already purified in 1926 and was proven to be a protein. Then in 1935, the last of the 20 natural amino acids was identified [35], thereby enabling determination of protein sequences. Just a few years later, the invention of electrophoresis by Arne Tiselius enabled separation and analysis of proteins in complex samples [36]. Subsequently in 1949, Linus Pauling *et al.* could show that sickle cell anemia is a molecular disease by detecting different types of hemoglobin between normal individuals and from individuals with sickle cell anemia [37]. They concluded that sickle cell anemia is a disorder where the genes are different between normal and diseased individuals, and by introducing the modification that causes sickle cell anemia in the genes; the hemoglobin molecules will be distinguishable. Finally in 1955, Frederick Sanger *et al.* determined the amino acid sequence of the first protein, insulin [38].

These findings came to influence the scientific method over the coming decades, in terms of making observations of proteins or phenotypes and then attempting to identify the corresponding genes to understand the mechanisms behind it. This approach resembled how Mendel studied the phenotype of peas and then hypothesized about the genes responsible for the phenotype. But unlike the 19th century when Mendel operated, by the 1970's the foundation for DNA sequencing was laid with the introduction of techniques such as Sanger

sequencing [39], this enabled researchers to study how changes in the DNA sequence could alter the protein of interest. The value of determining the genome only increased, and many modern techniques were introduced, allowing for numerous new discoveries. These included the creation of the first transgenic animal models to study proteins in 1981 [40], and the invention of the polymerase chain reaction in 1985 [41].

A new era in science emerged when the Human Genome Project (initially Human Genome Initiative) was launched in 1990 [42,43], and the term genomics was coined; referring to the large-scale analysis of the genome [44]. The project received billions of US dollars in public grants from the National Institutes of Health (among others) with the goal to map all the genetic material in the human body and the genomes of several model organisms within the next 15 years. In 2001, the Human Genome Project presented its first draft of the human genome, concurrent with another draft of the human genome from the privately funded project of Celera Corp. that started as late as in 1998 [45,46].

Since the start of the Human Genome Project, enormous sets of sequencing data have been generated. Less than halfway into the project, when some of the model organisms' genome had been mapped, it was clear that many genes were yet to characterized. For instance, for the commonly studied model organism *Caenorhabditis elegans*, half of the identified open reading frames in its genome had at that time unknown functions [43]. The result of the Human Genome Project would eventually lead to a new paradigm shift in biology, moving on from studies of known phenotypes of interest, to studies of genotypes without prior knowledge of function or phenotypic consequence [47]. The emerged field of genomics, would shift the working hypothesis of biology towards large-scale data generation rather than the classical hypothesis-driven biology [48], and in this process paving the way for many 'omics' fields such as transcriptomics and proteomics.

As the central dogma of biology states that information flows from DNA to mRNA to to protein, a direct result of the discovery of the genes with unknown functions by was that researchers focused on 'functional genomics' with e.g. animal model knockouts as proxies to fully understand the genome and the unknown genes [49]. Moreover, techniques started to emerge in the 1990's that allowed large-scale analysis of mRNA, thus enabling transcriptomics [50,51]. Large-scale analysis of proteins started to emerge and mature, traditionally

having relied on 2D-electrophoresis (from the 80's). Also, the replacement of Edman sequencing by mass-spectrometry as a tool for identification of proteins would further develop the field, and subsequently in 1994, the term proteomics was coined [52].

2. Proteomics

Proteomics, the field of large-scale analysis of proteins emerged in the shadow of the genomic revolution, with the promise of further understanding the function of all genes and ultimately complex biological systems [53,54]. Unlike genomics, the large-scale analysis of DNA and genes, the motivation for studying proteomics is that proteins are the class of molecules that perform most of the work in an organism, and not surprisingly, they are the group of molecules that most drugs target [55].

Do we really need proteomics?

A protein's function in an organism depends on many variables, including its localization, interactions with other molecules, structure, and abundance [53,56,57], thus making it difficult to infer the biological function in a perturbed system by measuring nucleic acids alone [58]. Also in many biologically important samples such as blood serum and plasma, the samples are cell-free and thus do not constitute any rich source of nucleic acids. Although it has been shown that circulating DNA [59] and RNA [60] exist in blood. Circulating DNA have been shown to be actively secreted into the blood by vesicles from healthy cells, through apoptotic bodies in cells that undergo cell death, but also directly leaked out into the blood by necrotic and cancer cells [61]. For the messenger RNA (mRNA), its presence in blood seems to be associated with different types of cancer [62,63]. However, circulating mRNA molecules are difficult to detect as they are present in very low abundance, possibly due to degradation by highly abundant RNases in the blood [60], and some reports have used up to 1 ml blood in order to analyze the mRNA [62,63]. Nonetheless the exact biological function of circulating DNA and RNA remains to be determined [61,64,65].

There have been several attempts in inferring the protein levels in cells and tissues by measuring the mRNA levels in steady state, although the results of those efforts have been inconclusive and debated in last decades [66-71]. In the

recent years there has been a breakthrough as the analytical tools have undergone major improvements for large-scale analysis of mRNA (RNA-sequencing) and proteins (mass-spectrometry). Interestingly, in one of these studies [68], the authors studied the mouse mitochondrion and compared the protein levels with the transcript levels of 527 protein-coding genes. The authors reported a correlation coefficient of 0.54 (Pearson's product moment) between their estimated protein and transcript abundances. Furthermore, by studying subsets of selected gene ontology annotations (GO), they could observe a very diverse correlation, e.g. protein-coding genes that were annotated for being in the ribosome (n=68) had a correlation coefficient (Pearson's) of 0, while for the 189 protein-coding genes that were annotated for "organelle inner membrane" showed a correlation coefficient (Pearson's) of 0.67. In addition, observations on enrichments of certain GO annotations and stability of mRNA and proteins have been done previously [70], implicating a functional attribute of regulation of the protein abundances. In addition, a recent publication by Edfors *et al.* [72] showed that by introducing a conversion factor between mRNA to protein, protein copy numbers per cell could be predicted in 9 human cells lines and 11 human tissues for 55 proteins. In another study, a different approach was used, first calculating ratios for each gene between mRNA and protein levels in 12 different human tissues, which then enabled prediction of protein expressions of thousands of genes [73]. As the understanding of the mechanisms for regulation of gene expression on mRNA and protein levels further evolves, in the near future it might be sufficient to only analyze the transcriptome and then infer the protein levels of cells or tissues in steady state. Whether these techniques can be expanded to apply for other proteomes, such as in body fluids (e.g. plasma and serum) remains to be seen, but is rather unlikely as the link between mRNA and protein production is lost as soon as the nucleic acid sequence leaves the cell. Until then, the benefits of studying proteins for inferring functions of genes of biological systems are indisputable as they are the "end-products", the effector macromolecules of an organism, and found ubiquitously, including in the body fluids, in all living creatures [52].

Protein structures

Proteins are the organisms' multi-tasker, capable of carrying out a wide variety of functions in our bodies, e.g. enzymatic catalysis, cell-to-cell communication, structural support, or even adapt into molecules that can target and destroy

harmful molecules. These incredible diverse molecular actions are dependent on the structures of different proteins [74,75].

All proteins are characterized by containing combinations of the 20 natural amino acids, where each amino acid consists of an amino group, carboxyl group and a unique side chain that determines its function [74]. These 20 natural amino acids are usually divided in three major groups, depending on the characteristics of the side chains: 1) hydrophobic, 2) polar and 3) charged side chains [76]. Furthermore, amino acids are linked together by forming peptide bonds, between the carboxyl-terminal (C-terminus) and amino-terminal (N-terminus), and thus amino acids that are linked together are referred to as peptides, whereas longer polypeptide chains are referred to as proteins.

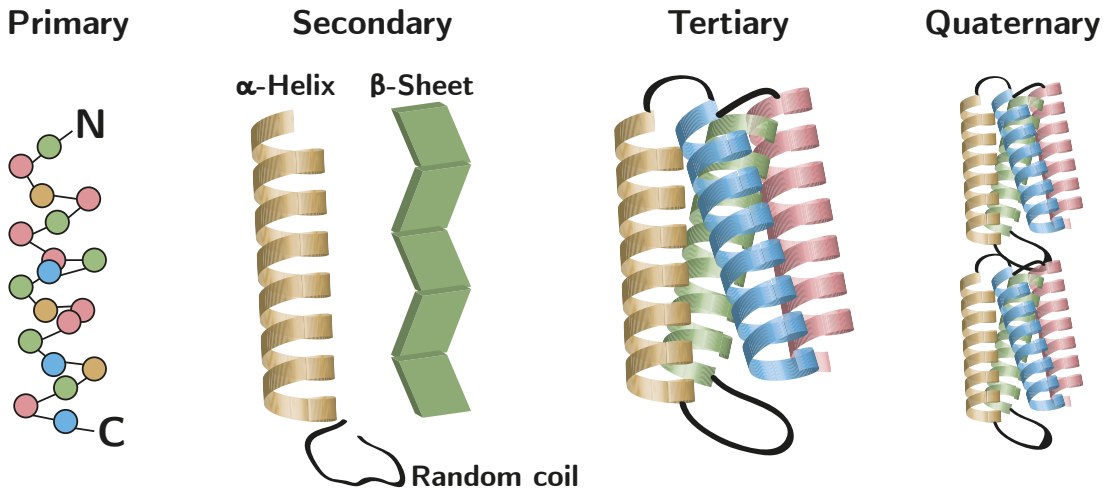


Figure 2. *The primary, secondary, tertiary and quaternary structure of a protein.*

Arrangements of the amino acids, both inter- and intra-protein are commonly described at four levels [75]. How the amino acids relate to each other in a one-dimensional space, and what the atomic composition of the protein is referred to as the primary structure. The structure of a protein is commonly described by writing the amino acid compositions from the N- to C-terminus. The secondary structure on the other hand is defined by how the primary structure of amino acids is assembled in a confined three-dimensional space. Furthermore, the secondary structure is mostly contributed by interactions in the peptide backbone, forming hydrogen bonds between the carbonyl- and amine-groups. Classifications and assignations of the different secondary structure types were introduced by Linus Pauling in 1951 [77], and the most common secondary structures are seen in Figure 2, i.e. helices, sheets and coils.

The tertiary structure is the structure that the protein adopts in a three-dimensional space and it describes how different entities, secondary structure, or elements in the protein interact with each other, e.g. by disulfide bonds. By arranging and combining two or more tertiary structures, the resulting fold will result in a quaternary structure.

Because of a shared evolutionary relationship between organisms, certain genes that are essential for life (e.g. energy metabolism) might be homologous and consequently will have a varying degree of sequence identity on a gene as well as on protein level. Therefore, there might be genes that have protein sequences that are almost completely identical, usually in the active site [78] and/or hydrophobic cores of e.g. globular proteins [79]. By studying sequence similarities between proteins that are homologous, researchers have been able to find out which residues that are important for the function of certain proteins [80]. This has enabled production of engineered proteins with advantageous features in various applications, e.g. improving antibody therapeutics [81], and in creation of more efficient subtilisins (protein-digesting enzymes) in detergents [82]. Sequence similarities have also been used to study functions of homologous proteins in model organisms to get clues of gene function [45,54]. In addition conserved amino acids, either due to catalytic function of the protein or structural function (i.e. if it is core residue or loop residue), have also a major implications in autoimmunity and epitope mapping, which will be further discussed in Chapter 4.

Studying the proteome

Altogether, there are 20,310 unique protein-coding genes (Ensembl release 89.38, Genome Reference Consortium Human Build 38 patch release 10) [83] in the human genome, representing just a fraction of our total DNA [45]. However, given that a protein-coding gene can produce different types of mRNA for the same gene, undergo alternative splicing, contain single amino acid polymorphisms and undergo posttranslational modifications (PTMs); the human proteome defined as all these different proteoforms, is estimated to be between 1-6 million different protein species [84]. However, for simplistic reasons one can regard the human proteome as the product of all protein-coding genes, also known as a gene-centric human proteome. Common methods to study the gene-centric proteome are roughly divided into two branches, mass-spectrometry-based methods and affinity-based methods.

Affinity-based proteomics

Methods for studying individual proteins have since a long time been available through e.g. enzyme-linked immunosorbent assay and western blot, and have been widely used to interrogate biological systems and functions of genes since their introduction in the 70's [85-87]. Despite the availability of these methods, it would take a long time before the first methods for large-scale analysis of proteins were developed, and this is primarily due to the lack of available antibodies [87]. As the generation of binders relies on information about the target protein, it was not until DNA sequencing methods began to create increasingly more complete drafts of the human genome that binders could systematically be generated. Consequently, as the availability of binders increased through large-scale initiatives such as Affinomics and the Human Protein Atlas project in the last decade [87-89], there has been a plethora of different methods for large-scale analysis, including high throughput and/or high multiplexing capacity, for studying the proteome. As the cDNA microarrays emerged in the mid 90's and were used for studying the transcriptome, equivalent microarrays for analysis of the proteome emerged a few years later. Initially, most of these microarrays were used to study protein-protein interactions [90-92] or interactions with other molecules, e.g. carbohydrates [93,94] and enzyme-substrates [95], but eventually antibody microarrays were made [96]. The concept and basics of these antibody arrays were already laid in the 80's by Ekins *et al.*, who suggested that miniaturization of analyte-binding assays (e.g. immunoassays) increases sensitivity [97]. Since then, many different types of microarrays have been developed for large-scale analysis of proteins, including antibodies immobilized on planar surfaces [96,98,99], antibodies on suspension beads arrays [100,101], but also other binder molecules immobilized on solid surfaces [102,103]. Furthermore, other proteomic methods have been employed that uses antibodies (and other binder molecules) including: a protein-detection method called proximity extension assay to study proteins in blood [104], a recent article that used antibodies for immuno-fluorescence in cells in order to create a subcellular map of 12,003 human protein-coding genes [57]. In another publication antibodies towards 16,975 human protein-coding genes were used to map 44 different human tissues and organs by immunohistochemistry [105].

Human Protein Atlas project

One of the previously mentioned large-scale initiatives aiming at producing affinity reagents is the Human Protein Atlas (HPA) project. The HPA project was launched in 2003 and involves high-throughput generation of antibodies targeting all human proteins, as defined by the protein-coding genome. Thereby enabling the whole proteome to be studied and to produce proteome-wide atlases of the proteins in tissues and cells [106], which have resulted in a comprehensive map of the human tissue proteome [105] as well as the subcellular proteome [57]. Furthermore, all data that is generated within the project is published on a publicly available database, www.proteinatlas.org, which as of July 2017 hosts proteome analyses based on 25,682 antibodies targeting 16,998 unique proteins.

As previously described [106-109], antibodies within HPA are created by first generating protein fragments, so-called protein epitope signature tags (PrESTs). These PrESTs are in general ranging from 50-150 amino acids with a median and mean of around 80 amino acids. They are designed *in silico* to represent a unique protein-coding gene by requiring that each 50 amino acid window in the PrEST should have less than 60% sequence identity to other human proteins, and each 10 amino acid window should have a maximum sequence identity of 80%. Each PrEST is then cloned into *Escherichia coli* cells, expressed and then purified, before it is used to immunize rabbits. Resulting antiserum is subsequently purified on columns carrying PrESTs. The quality of the purified polyclonal antibody is further tested: first on western blot probed with lysates from cells over-expressing full-length protein, as well as tissue samples and plasma. Then the antibody is tested on protein arrays consisting of 383 randomly selected PrESTs as well as one PrEST specific for the polyclonal antibody is created. Antibodies that pass these quality tests will then be used for subcellular profiling [57] and tissue profiling [105] that will be published on the database, Figure 3.

Also, PrESTs generated in conjunction with the project have also been utilized for studying potential autoantibodies [110,111], which will be further discussed in Present Investigations.

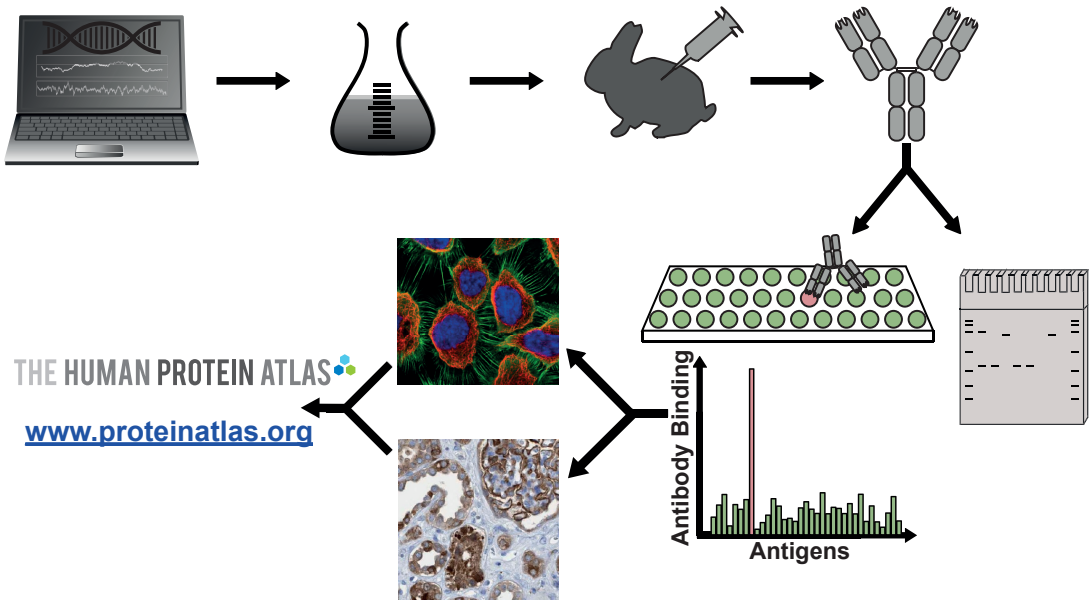


Figure 3. *Antibodies are systematically generated in the Human Protein Atlas workflow, first (upper left) by designing a suitable PrEST sequence through bioinformatics, which is then cloned and then expressed in E. coli. The purified PrEST is then used to immunize rabbits and the resulting antisera is purified and tested on protein arrays and western blot to determine its selectivity. Each passed antibody is used to map the spatial distribution of proteins in 44 different human tissues and organs, and in at least three different human cell lines. Annotated protein expressions are then published in a publicly available and interactive database.*

Mass-spectrometry-based methods

As mentioned in the previous chapter, mass spectrometry (MS)-based proteomics methods gained popularity among researchers in the early 90's as it provided higher sensitivity and higher throughput, and could thereby replace its progenitor Edman degradation [112]. The shift can be attributed to the innovation of two different ionization methods, matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI), and to this day, these ion sources still remain the foundation of all MS-based applications for identification and quantification of proteins [113]. Since the introduction in the early 90's, a lot of technical advancements have been seen in the field of MS-based proteomics [114-118], including increased ionization efficiency, improved mass-analyzers as well as fragmentation techniques, but also to different sample preparations techniques [119,120] and bioinformatics analysis [121 ,122,123]. The identification of proteins by MS-based methods ultimately relies on measuring mass to charge ratios of ions very selectively and with accuracy. This gives MS-based methods an advantage over affinity-based

methods as it also can identify and measure proteins that have undergone post-translational modifications in a proteome-wide manner. However, there are still challenges in studying PTMs with MS-based methods, including pinpointing the correct site of PTMs and making such analysis sensitive [124]. Also, the introduction of isotope-labeled standards allow for precise quantification of proteins, especially in combination with targeted approaches such as multiple reaction as shown by Chambers *et al.* in 2015, where 97 proteins were quantified in sample material taken from dried blood spots [125]. Another quantification approach, based on label-free absolute quantification was published last year [126], where one microliter human plasma provided enough material to identify and quantify 284 proteins in a reproducible manner across 15 individuals. As these individuals were sampled longitudinally, the authors could show the intra-individual variability for the quantified proteins, thereby opening up new dimensions in understanding proteome changes over time. Furthermore, in an article published in 2014 [73], Wilhelm *et al.* presented the first draft of the human proteome from data collected from 16,857 MS experiments, including various tissues and cell lines. This revealed identification of 18,097 proteins and their distribution in human tissues and cell lines.

Challenges in studying the human proteome

In a recent study by Uhlén *et al.* [105], the authors suggest that there are in fact at least 3100 genes in the human proteome that lack evidence of existence at protein level. Some of these are suggested to be falsely annotated as protein-coding genes. Although not discussed in that study, another reason for this can be attributed to the variation of the protein abundances in the human proteome and that those genes thereby go undetected, as suggested by Kim *et al.* [127]. In the study by Uhlén *et al.* [105], transcriptional expression levels in 32 different tissues and organs for all the human protein-coding genes. A total of 44% of all protein-coding genes were expressed across all tissues, and transcripts corresponding to 12% of all the protein-coding genes were enriched in any of the 32 studied tissues. Given the relationship between the transcripts as proxy for protein expression (as previously discussed), these findings consequently suggest a tremendous heterogeneity in the proteome between different human tissues. The ultimate aim of proteomics is to understand the function of protein-coding genes and thereby understand biological systems. Since proteins have different structures, can interact with other proteins, are

compartmentalized to different locations in cells and tissues as well as vary in abundance, hence a variety of methods have been developed in the last decades to enable their detection [53,54,128].

Although we have seen that the transcript levels of all the protein-coding genes vary from tissue-to-tissue and that they can differ from gene-to-gene (with the similar findings on a near to proteome-wide coverage at the protein level) [105,127]; the complexity and dynamic range of the human plasma proteome is even greater than for the tissue- and cellular proteomes [129]. Unlike when studying nucleic acids where it is possible to amplify the molecules through PCR, there is no method for amplifying the proteins. Therefore, one of the greatest challenges has been to develop methods that are sensitive enough to detect the proteins representing 20,310 unique protein-coding genes as they vary in abundances [54,128,130]. Due to large dynamic range of proteins present in many biological sample sources, including cells and commonly used body fluids (e.g. plasma and serum), there could be differences of as much as 10 orders of magnitude between the highly abundant plasma protein albumin and the low abundant cytokine interleukin-6 [129]. This dynamic range can be compared to, having the entire population of humans in one location and then trying to find a single person of interest [131].

Given that disease-relevant, thereby also clinically relevant proteins are often found at much lower plasma concentrations than abundant plasma proteins such as albumin, due to e.g. tissue leakage and/or abnormal secretion from tumor or cancer tissue [131], there is a need to be able to analyze lower abundant proteins [131]. Therefore, there have been tremendous efforts in improving the methods for analyzing the proteome, which has been highlighted in several recent studies [126,129,132,133]. Mass-spectrometry-based proteomics has for a long time been limited in sensitivity mainly due to limitations in sample preparation and technological limitations in the mass spectrometer itself (e.g. ionization and peptide separation). This is illustrated in a paper from 2004, where three different MS-methods, including a literature search, were compared in their capability of identifying human plasma proteins [134]. Out of 1175 unique proteins that were found in any of the the four methods, only 4% showed an overlap between all methods, and only 17% of the proteins were found in more than one method [134]. Additionally, in a paper published in 2009 [135], 27 laboratories were given a blinded sample consisting of a mix of 20 human proteins, but only 7 labs were able to successfully detect all 20 proteins. Since 2004, sample preparation techniques has evolved and have

become standardized [120,126,136], including better liquid chromatography systems used for upstream peptide separation. This, together higher resolving power and faster scanning rates of mass spectrometers have enabled more sensitive peptide identifications and thus protein identifications [137].

For affinity-based proteomics, the limitations in creating sensitive assays have been due to the lack of availability of good protein-binders. As the availability of binders has increased through large-scale initiatives [88,89], including recombinant binders [87], the focus has been on creating more sensitive read-out techniques. One of these methods is proximity extension assay PEA[104], similar to proximity ligation assay, which is based on conjugation of affinity reagents with at least two different specificities per protein, with nucleic acids. Here, the detection of proteins is based on proximity of the antibodies when binding to a target protein, which enables the nucleic acids to hybridize, followed by amplification and read-out through PCR. By detecting nucleic acids it is possible to amplify the signal through PCR and thereby achieve very high sensitivity [88]. Also commercial methods such as the Erenna Immunoassay System can achieve very sensitive read-outs by minimizing the background through measuring each antibody-protein complex individually in solution, with reported limit of detection between 0.01-0.12 pg/ml for ten different proteins [138].

Although sensitivity is an important feature for any proteomic method, specificity is at least an equally important, since a sensitive proteomic method that has a low specificity can in theory bind anything [87,139,140]. This is a huge challenge for researchers, specially within the field of affinity proteomics, since the identification of proteins are, unlike when using mass-spectrometry, completely blind, i.e. there is no way to know by the test itself that the correct protein is measured [139]. As extensively discussed in several recent articles [133,139], the lack of specificity or presence of cross-reactivity is a major problem for the experiments and especially for assays that rely on a single binding event of an antibody or similar. Therefore, there has been a lot of focus in creating methods that are more specific [140], e.g. through sandwich immunoassays, which relies on the detection of two binders with different specificity thus decreasing the chance of cross-reactivity in the two independent events. In a similar way with PEA (and PLA) that relies on proximity of multiple binders in order for the nucleotides to hybridize, which is minimizing the risk of off-target detection [104]. However, multiplex methods that use multiple antibodies need to be thoroughly validated as the risk for cross talk

increase quadratically with the number of targets analyzed (if two antibodies are used per target) [141], thus making conventional multiplex immunoassays very time consuming to design without cross-reactivities. Nevertheless, similar methods have been developed with the potential of increasing specificity of an assay: sequential multiplex analyte capturing [142] and dual capture assays [143], which use two sequential single capture events, by creating pseudo-sandwich immunoassay in time rather than in space (as conventional sandwich immunoassays) and thus avoiding the need for developing antibody pairs that do not cross-talk. Finally, there is a proposal [133] written by an international working group for antibody validation that formulated five different methods to validate the specificity of an antibody; genetic methods (e.g. mRNA knock-downs), orthogonal methods (e.g. MS-based method), independent antibody, tagged protein expression (e.g. GFP-tagged protein co-localizes with antibody), immuno-MS (e.g. captured proteins from an antibody). These criteria will very likely form the basis for all future antibody validation, whether the antibodies are used for proteomic studies or not.

In analogy to the affinity proteomics, the specificity for MS-based proteomics is ultimately determined by how well the mass over charge ratios can be measured and how successful the resulting fragment spectra can be mapped to protein sequences by database search algorithms, and ultimately assign them to proteins [144]. During the last decade of development of the MS-based methods, both analytical improvements such as better mass-filters, ion fragmentation techniques, but also improvements in bioinformatics have made the MS-based methods the ultimate tool for specific measurements of complex protein samples in comparison to affinity-based proteomics [121,145,146].

3. Antibodies

The immune system is an organized cellular machinery present in almost all forms of life, whose role is to protect the organism against foreign organisms and molecules. In jawed vertebrates the immune system can be divided into two branches, the innate and the adaptive immune system, which have distinct functions and roles but can activate each other. As the name suggests, the innate immune system, is the first line of defense against pathogens and relies on recognition and degradation of evolutionary conserved patterns, as in damage-associated molecular patterns and pathogen-associated molecular patterns. In contrast, the adaptive immune system requires days to fully respond to pathogens, as it relies on learning to what to recognize and degrade. Also, the magnitude of the response in the adaptive immune system is far greater than the innate response. It will on subsequent exposure to a pathogen become even more rapid and effective. In addition, the adaptive immune system can be classified into two groups; cellular or humoral immunity. Whereas the former is mediated by T-lymphocytes, humoral immunity is mediated by B-lymphocytes and their soluble products, antibodies, also known as immunoglobulins.

Structure

Although the existence of antibodies was demonstrated already in the late 19th century by Kitasato Shibasaburō and Emil von Behring [147], the antibody structure with its chains and domains remained unexplored until 1959 when Rodney Porter could successfully show that antibodies can be split into two types of fragments [148]. These fragments were later named Fragment antibody binding (Fab) and Fragment crystallizible (Fc).

Today, we also know an antibody consists of two identical light chains and two identical heavy chains that are all linked together by at least three disulfide bonds. In addition to the Fab- and Fc-fragments, the antibody contains other important regions such as the hinge region, which enables the Fab parts to be

flexible and the J chain, which is present in some antibodies, linking together antibody monomers.

Furthermore, each heavy chain contains four or five immunoglobulin domains, whereas the light chain contains two domains [149]. The N-terminal immunoglobulin domain of each chain is referred to as the variable domain, due to the high variability in amino acid composition compared to antibodies from the same species. Likewise, the remaining domains are referred to as constant domains, due to their low amino acid variability [74]. One of the main roles of the constant domains is to carry out effector functions, such as activation of the complement system or binding Fc receptors on cells [150]. The outermost region of the variable domain is referred to as the complementary-determining regions (CDRs), or the hypervariable loops. These CDRs also constitute the majority of the antigen-binding surface and since they are the most variable regions of the antibody they make up most of the diversity seen in antibodies. The light chain of the antibody can contain two different isotypes, kappa chain or lambda chain. Although, antibodies containing kappa chain is more prevalent in our bodies [149,151], so far no functional differences for the antibody related to the isotypes have been observed [74].

Isotypes of the antibody heavy chain is on the other hand significant for the function of the antibody, and determines its class. There are five different heavy chain isotypes, the alpha, delta, epsilon, gamma and mu heavy chains, which determines the five antibody classes, namely IgA, IgD, IgE, IgG and IgM. Moreover, the heavy chain of the alpha and gamma isotypes can be divided into two and four sub-isotypes respectively. Different antibody isotypes, have different roles and functions in the human body and certain isotypes are preferred by certain tissues, for instance: IgE is known for being involved in allergy and reacts to allergens on the skin or in the lungs and IgM is the first line of defense in blood and involved in activation of the complement system [150,152]. In recent years, it has been shown that different isotypes affect affinity and specificity of the antibody, and ultimately its activity [153]. Therefore, the constant region does not only carry out effector functions, but also affect the specificity of the antibody through structural changes [154].

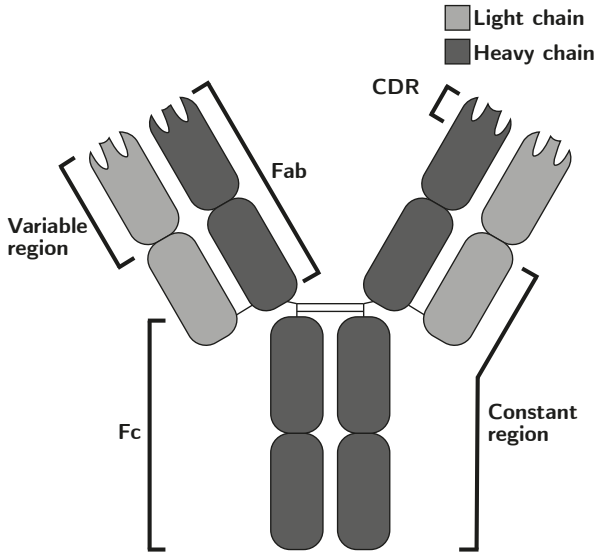


Figure 4*. Schematic overview of the antibody (immunoglobulin G) structure showing the CDRs, complemtary-determining regions, Fab, fragment antigen binding, Fc, fragment crystallizable, constant and variable regions, and light and heavy chains.*

Function

How does the antibody bind its target? Historically scientists hypothesized on antibodies bind their targets, antigens, through a "lock-and-key" mechanism, as introduced in in the late 19th century. Today we know that the antibody can also undergo conformational changes upon binding to an antigen. Such changes are possible due to the degree of freedom that is enabled by the flexible hinge region of the antibody, as well as the nature of the antibody-binding site, as explained by Koshland's induced fit mechanism [155].

Moreover, the binding strength of an interaction between an antibody's binding site and its antigen is referred to as affinity, and it depends on the composition of the interfacing surfaces of both entities, i.e. antibody and antigen. Affinity of an antibody is commonly estimated by determining the dissociation constant, K_D, of the reaction between an antibody and its antigen subsequently forming an Antibody:Antigen complex, which can be exemplified as following equilibrium (eq. 1):

$$\text{Antibody} + \text{Antigen} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Antibody:Antigen} \qquad \text{(eq. 1)}$$

From eq.1, the dissociation constant, K_D, can be defined as the ratio between the off-rate and the on-rate, which can also be expressed as the ratio between the concentration of the reactants and the concentration of the product (eq. 2):

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[\text{Antibody}]\,[\text{Antigen}]}{[\text{Antibody:Antigen}]} \quad \text{(eq. 2)}$$

Note that eq.2 is only valid for monovalent binding of the antibodies. Multivalent binding, meaning that more than one antigen can bind to the antibody, may however occur. For example, IgG can bind two antigens, one per each Fab region, and IgM can form pentamers and therefore bind up to 10 antigens simultaneously [149]. Therefore, the rate of binding, affinity, of an antibody:antigen interaction can be affected by the antibody's valency [156]. Hence, avidity is the total binding strength gained by valency through effects such as forced proximity [157]. Therefore, an antibody with a low intrinsic affinity can achieve high functional affinity due to avidity, through its multivalent capacity [157-159]. Due to the increased functional affinity of an antibody (or any other binder) by multivalency, there have been considerations [160,161] about designing an antibody, e.g. towards a protein expressed highly on the surface of cancer cells, and lower levels in normal cells. An antibody with high intrinsic affinity will bind normal cells to a higher extent, and will be less probable for it to migrate to the site of interest (i.e. where the cancer is located), compared to an antibody with lower intrinsic affinity. Multivalent antibodies may therefore be beneficial for therapeutic antibody, due to increased functional affinity and longer retention at the surface of cancer cells.

As discussed above, an essential point in order to understand the function of an antibody is to determine its affinity for its target and other kinetic constants. The first paper that describes the kinetics behind such interaction involving an antibody was published in 1946 [162]. Equilibrium dialysis [162] was then used to measure free antigen and antibody-conjugated antigen to enable reaction constants to be determined. Since then, several innovations have enabled faster and more accurate determinations of kinetic constants. These inventions include the methods surface plasmon resonance, i.e. an optical technique that enables label-free detection of the interaction between antibody and antigen

(among others) [163]. Other methods for measuring reaction kinetics include further development of enzyme-linked immunosorbent assays (ELISAs) [164] and the use of thermophoresis for analyzing protein-antigen interactions [165].

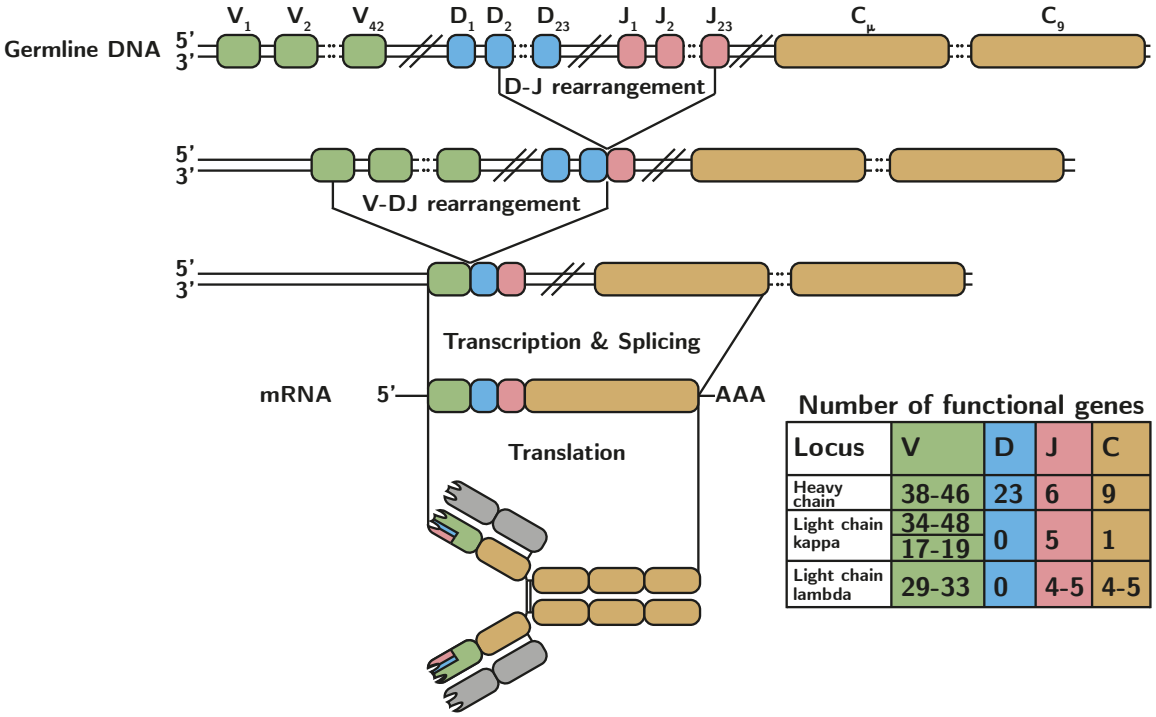
The underlying molecular forces that determine the affinity, and ultimately keep the antibody and antigen together have different characteristics. A common denominator is that all these interactions are weak and each has only a equivalent binding strength that is less than 10% of a typical carbon-carbon covalent bond [166]. The type of interactions and molecular forces that take place between an antibody and antigen include electrostatic forces, hydrogen forces, hydrophobic effects and van der Waals forces [167], where the main contributor are the hydrophobic forces [166]. These forces originate from non-polar amino acids of the antibody and antigen. When the antibody and antigen are not in complex, they have highly ordered water molecules interacting with the non-polar amino acid residues. Therefore, it is energetically favorable for the water to be displaced in the interface of the antibody and antigen when they are in proximity and instead contribute to their binding hydrophobic interactions [76,168].

Achieving Diversity of Binding/Recognition

Antibodies are produced by our immune system, and specifically by the B-cells, to protect us against pathogenic and harmful entities. Therefore, it is crucial that antibodies have the ability to recognize as many potential molecules as possible that it might encounter in the future. But how can the antibodies recognize molecules it has not yet encountered? And how can antibodies have a seemingly endless specificity from only a finite set of genetic information?

One of the key mechanisms behind creating this diverse repertoire of antibodies derive from the fact that the amino acid sequence of an antibody is not encoded by a single gene, but instead by recombination of different gene segments, which eventually build up the information to generate the antibody. This fundamental insight of the rearrangement of the genes of immunoglobulins was made as late as in 1976, for which Susumu Tonegawa was awarded the Nobel Prize [149,169]. Since then additional mechanisms for generating antibody diversity have been discovered [149].

As discussed in the introductory section of this chapter, the antibody consist of two different chains, the heavy and the light chain, The two light chains (kappa and lambda) are encoded by gene families located on different chromosomes and the heavy chain genes (alpha, delta, epsilon, gamma or mu) are found yet another chromosome. The light chains are encoded by three gene segments in the germline genome, the variable (V) gene segment, the joining (J) gene segment and the constant (C) gene segment, while the heavy chains are encoded by a V, J, C and diversity (D) gene segments in the germline genome [149]. Each V, D, J and C gene locus contain several different gene segments, as can be seen in Figure 5, where e.g., the human heavy chain gene loci contain 38-46 different V genes, 6 J genes, and 9 C genes [170]. By combinatiorial effects of the gene segments in the heavy chain and the light chain loci, (38 x 23 x 6) x ((34 x 5) + (17 x 5) + (29 x 4)) ≈ 2 million different combinations of antibodies can be produced, Figure 5.



Locus	V	D	J	C
Heavy chain	38-46	23	6	9
Light chain kappa	34-48 17-19	0	5	1
Light chain lambda	29-33	0	4-5	4-5

Figure 5. *Schematic overview of somatic recombination of immunoglobulin genes and the number of functional genes per locus and gene segment resulting in (38 x 23 x 6) x ((34 x 5) + (17 x 5) + (29 x 4)) ≈ 2 million potential combinations.*

As described by Owen *et al.* [149], during the B-cell development in the bone marrow, each B-cell undergoes somatic recombination of genes to create the variable region of the future immunoglobulin. The process is carried out by several different proteins, named V(D)J recombinase [167], whose roles can be summarized as breaking, repairing and ligating DNA strands, until a complete

antibody gene is created containing one element of each of V, D and J gene segments (Figure 5). The resulting mature B-cell that have undergone the somatic recombination will eventually express IgM and IgD receptors on the cell surface, ready to encounter antigens.

Apart from the recombination of gene segments, additional diversity is created during this process throughnuclease activity by V(D)J recombinase, leading both to loss of nucleotides and addition of nucleotides due to assymetric cleaveage of DNA strands, and random addition of nucleotides by terminal deoxynucleotidyl transferase.

Additionally, an antibody can undergo mutational processes, called somatic hypermutation (SHM). There the diversity is further increased, which takes place in follicular B-cells during a T-cell-dependant B-cell response. After the B-cells have maturated in the bone marrow, they circulate in the blood and the lympatic vessels. Once in a while, they migrate to the lymphoid follicles for survival stimuli. In case a folicular B-cell encounters and binds to an antigen by its membrane-bound antibodies, the membrane-bound antigens gets internalized in the B-cell through endocytosis. There it is processed and presented on the major histocompatibility complex (MHC) class II on the B-cell surface.

If the B-cell is in the secondary lymphoid tissues, the B-cell migrates to the border of T-cell-rich zone and B-cell zone, and subsequently interacts with activated T-helper cells. T-helper cells bind through their T-cell receptors to the MHC class II receptor on the B-cells, and in addition give co-stimulatory responses to the B-cell through signaling of CD40L of the T-cell to CD40 on the B-cell, and binding from the CD28 receptor on the T-cell to CD80 and CD86 on the B-cell [149]. This leads to secretion of cytokines from the T-cell that activates and induces proliferation the B-cell. Some of the activated B-cells migrate to medullary chords where they differentiate into plasmablasts, while other B-cells migrate to the interfollicular region in the secondary lymphoid tissues were they form germinal centers (GCs) [171]. GCs are sites in the lymph nodes where specific follicular helper T-cells stimulate the B-cells to clonally expand and go through SHM, but also class switch recombination (CSR). Although CSR is the process of changing the constant region of the heavy chain resulting in isotype change of the antibody, this is not regarded as a mechanism of increasing the diversity of the antibody [149]. However, it can be argued that isotype change of an antibody are means of the B-cell to further

diversify by alterination of its effector function [150]. It has recently been suggested by Casadevall & Janda [153] to also affect the binding strength (and specificity).

Nevertheless, SHM adds diversity to the variable region of the antibody by introducing mutations to the light and heavy immunoglobulin genes through activation-induced cytidine deaminase (AID). This protein together with other proteins involved in SHM increase the frequency of mutations by six orders of magnitude compared to the basal mutation rate in the genome [172]. During transcription, AID binds single-stranded DNA and preferentially targets cytidine residues of the DNA and deaminates them to uridine. The resulting double-stranded DNA then contains a mismatch that in turn activates different error-prone repair mechanisms resulting in somatic mutations of one or more bases [149,172]. The mutated B-cell clones then compete with each other for binding of a limited amount of antigen, which is required for receiving proliferation and survival signals mediated by the follicular T-helper cells. This is mediated by interaction between the MHC class II on the B-cell and the T-cell receptor. This process, known as affinity maturation will result in loss of the low-affinity B-cell clones, while the high-affinity B-cell clones will survive and mature, thus create a more refined and diversified panel of antibodies [149].

Taking all the above-mentioned mechanisms for creating antibody diversity together, the number of antibodies with potentially different specificities is tremendous. Estimations of the potential pre-immune antibody repertoire, excluding SHM, point at more than 10^{16} different antibodies [150], thus demonstrating how the antibody repertoire can cover seemingly limitless range of specificities.

Antibodies – A toolbox for molecular biology

The immense capability of antibodies to recognize diverse molecules was realized by scientists long before the underlying mechanisms that enabled such characteristics were known, and have since then been exploited to study biology. Many regard Louis Pasteur as one of the founders of the field of immunology [149]. In 1881 he could show that immunity can be induced by challenging to an avirulent form of a pathogen [173,174]. This discovery paved the way for Charrin and Roger's work in 1889, where they studied rabbits inoculated with the bacterium *Pseudomonas aeruginosa* [175]. Following the

inoculation, they could show that there is a substance present in the blood of rabbits that interacted with the bacteria. Further characterization of this substance was made a year later by Kitasato Shibasaburō and Emil von Behring [147]. By studying mice inoculated with the tetanus-causing bacterium followed by injections of cell-free component of blood of previously inoculated mice, they could conclude that there was a substance in the cell-free part of blood that could protect organisms from toxins, which was named antitoxins and later antibodies [176].

These above-mentioned experiments by Pasteur, Charrin, Roger, Shibasaburō and von Behring had a great impact and paved the way for many scientists in the coming years [177]. Eventually scientists started to realize the potential that antibodies possess in binding specific molecules. This resulted in the first diagnostic test for typhoid fever and for cholera on infected animals, as presented by Gruber and Durham in 1896 [178]. Their work was followed by the implementation of their method for diagnostic purposes of other diseases in humans [179,180]. Most notable is the work by Fernand Widal that would later name these diagnostic tests, i.e. Widal's test [181,182]. Throughout the 20th century antibodies later became widely known for detection and identification of other molecules. Methods using antibodies in blood (antisera), were further developed and used for several groundbreaking applications [183], including the discovery of DNA [28,30].

In the coming decades, scientists started to analyze the composition of blood and antisera [184] with the introduction of techniques such as electrophoresis [36] and even better with the protein sequencing techniques [38]. This enhanced the understanding of the nature of the antibodies. Furthermore, methods such as enzyme-linked immunosorbent assay (ELISA) [85], radioimmunoassay (RIA) [185] and a sensitive method of immunohistochemistry (IHC) [186] were developed. These techniques enabled the use of antibodies to perform sensitive analysis of proteins, both in solution and for tissue analysis. Around this time, in the 1960's, concerns were raised regarding the specificity of the antibodies in the antisera due to the potential cross-reactions of the antibodies in the antisera [183,187-189]. Subsequently new techniques, such as affinity chromatographic methods, enabled antigen-specific antibodies to be purified from antisera, hence resulting in improved research reagents [190]. Antibodies are referred to as polyclonal antibodies when produced by multiple B-cell clones. They are commonly purified either on their antigen of interest or

protein A/G, which are bacterial cell wall proteins with affinity to antibodies [167,191,192].

In the mid 1970s a revolutionary method was invented that enabled fusion of spleen cells from an immunized animal with myeloma cells [193], to produce hybridoma cells that are both capable of growth *in vitro* and secrete antibodies. This in turn made it possible to produce antibodies derived from a single B-cell clones, hence monoclonal antibodies. Although this was not the first attempt in creating monoclonal antibodies [194,195], it was by far the most successful and is still being widely used today. In contrast to the polyclonal antibodies that are limited to the amounts of antibody a host animal produces, the monoclonal antibodies are a renewable source and can theoretically be produced in unlimited amounts. However, although the hybridoma technology has been improved, the process is still much more complex than generation of polyclonal binders. This has a impact on the time and cost and makes polyclonal antibodies both cheaper and faster to produce compared to monoclonal antibodies [196,197].

What are the differences between antibody preparations of different clonality? In contrast to monoclonal antibodies, polyclonal antibodies consist of antibodies produced by multiple B-cell clones, thus cover different specificities and bind to multiple sites on an antigen. Even though there are apparent differences in the potential specificities between the two different preparations, a recent article [133] suggests that the performance of an antibody can only be verified for the specific application and context for which it is intended to be used. Therefore each antibody preparation needs to be evaluated individually, based on the intended use. A common argument and what some immunology textbooks propose [149,167,198], is that polyclonal antibodies are by default more cross-reactive or prone to result in non-specific binding [199] than monoclonal antibodies. These arguments are, as it seems incomplete [200] since the antibodies need to be judged individually [133]. The thought that polyclonal antibodies are more cross-reactive seem to derive from the time when polyclonal antibodies were prepared as antiserum, and thus contained various specificities that were not related to the immunogen [198], and thereby causing a higher apparent cross-reactivity [108]. But as discussed previously; since the entry of the affinity chromatography, antibodies in antisera can routinely be purified for each antigen before use [197]. Although polyclonal and monoclonal antibodies have been designed to bind a certain antigen, they are all polyspecific [200] and will both have the possibility of off-target interactions.

An antibody can both bind similar structures to the immunogen that are present in other proteins, but also interact with other, non-similar structures. Furthermore, a polyclonal antibody will have a more heterogeneous polyspecific off-target interactions than a monoclonal which will have less polyspecific off-target interactions [201]. In theory, a polyclonal antibody preparation might not be as sensitive for off-target interactions as a monoclonal antibody, since the off-targets will be diluted in comparison to the on-target interactions [200,201].

The clonality of polyclonal antibodies can be advantageous since it allows recognition of multiple antigen sites and thus it might be more successful than monoclonal antibodies in methods in which the antigen is denatured (e.g. Western blot, immunohistochemistry, etc.) [197,202]. Also, polyclonal antibodies are advantageous for the detection of antigens that are heterogeneous, e.g. a protein in blood with different post-translational modifications. The same feature may however also be viewed as a disadvantage in case only a single form of the protein is wanted for binding.

Another important point in the comparison between monoclonal and polyclonal antibodies is the binding strength. Since polyclonal antibody preparations contain diverse sets of antibodies with different binding strengths, their average affinity is usually lower than for monoclonal antibodies [197]. For monoclonal antibodies, B-cell clones producing high-affinity antibodies can specifically be selected through affinity determination and selection methods, thereby achieving higher affinity than in polyclonal antibodies [149,203].

Also in the last decades, other binder molecules [204,205] have been developed with the advantage of not relying on immunization of animals during the production. Many of these binders were thought to solve the problems with lack of specificity and reproducibility of antibodies [109,206-210], but so far with limited success as these binders face similar challenges as antibodies [133,210]. Some of these binders are based on the antibody structure e.g. Fab fragments, single chain fragment variable antibodies where the variable domain of the heavy and light chains are linked with a peptide, and single variable domain binders. There are also several examples of binders that are not based on the antibody structure, such as monobodies that are based on the fibronectin protein fold, DARPins that are based on the ankyrin fold, whereas aptamers are based on nucleic acids and affibodies are based on protein A from *Staphylococcus aureus* [103,205]. As these recombinant binders do not undergo

any *in vivo* affinity maturation and selection, display methods such as phage display and bacterial display are commonly used to enable screening of large gene libraries to identify and select sufficient binders [205].

In conclusion, there is a plethora of different protein binders produced both naturally and recombinantly that have been used to study biology, and all with their certain advantages and disadvantages over the other. As more binders are produced today, it is very important that the research community, together with the producers, validate their binders in order to ensure reproducibility and specificity, e.g. by following the guidelines of the International Working Group for Antibody Validation [133].

4. Antigens and Epitopes

As every protein in our body has its defined specificity towards certain molecules, and sometimes also has less specific interactions with other molecules, it is of great interest to explore the characteristics of such interactions. This is especially true for antibodies since this class of molecule is, as discussed in the previous chapters, widely used to study the proteome. But they are also used to study immune reactions, e.g. towards a microbe or cancer tissue, and used in therapy including antibody therapy and vaccines, among others. In this chapter, the interaction between antibody and protein will be discussed, how these can be studied, as well as the implications that certain interactions can have for diseases.

Epitopes and Paratopes

As discussed in chapter 3, antibodies interact with their antigens through binding to their complementarity-determining regions. These CDRs are found in the loops between the beta-barrels, with three regions per chain, thus creating an antibody binding-pocket consisting of 6 CDRs. Interestingly, improvements in affinity by somatic hypermutation mainly affect the amino acid compositions of these loops [172]. Also, the variability for the heavy-chain CDR region is higher than the light-chain CDR [167]. This is due to the somatic recombination processes, where the variable domain of the heavy chain has an additional gene segment compared to the variable domain of the light chain [167]. In addition, a recent study made on crystal-structures of 140 antibody-antigen structures have identified that the number of residues in contact with target molecule is higher in the heavy chain than the light chain [211]. These contact residues of the antibody are referred to as the paratope, while the contact residues on the antigen side are referred to as the epitope.

Furthermore, several recent studies has shown that hydrophobic amino acids, such as alanine, valine and leucine are found to a lower frequency in the epitope residues relative to non-epitope regions [212-214], implying that it

might be related to the surface exposure as hydrophobic residues are found less frequently on the surface [76]. It should be noted that the presence of these enrichments in the epitopes are not clear, as there are conflicting results in a recent study that reported no significant differences between epitope residues and surface exposed residues of the antigen [215]. In a related study [216], antibody responses towards amino acid compositions of 12,634 immunogens were studied. There it was found that hydrophilic residues in the immunogen cause stronger antibody response than hydrophobic residues.

Also, studies have been performed on paratopes that show tryptophan and tyrosine to be enriched [213,214,217-219] compared to the average amino acid composition of an antibody. Interestingly, the most frequent amino acids in the paratopes are found at higher frequency than the most commonly occurring amino acids in epitopes [212,214]. This suggests a higher structural conservation in the paratopes compared to epitopes. It has also been shown that tyrosine paratope residues interact to a higher extent with hydrophobic and charged amino acids in epitope, compared to other amino acids [218]. This is potentially due to its ability to mediate different types forces in the binding [201,215]. In addition, certain amino acids are found in different frequencies in the six CDR. This suggests that the residues in the paratope prefer to bind different types of amino acids in epitopes and thus making the epitope indistinguishable from the amino acid composition of protein surfaces [215].

Epitopes can be divided into two groups: continuous epitopes, also known as linear epitopes or sequential epitopes; and discontinuous epitopes, also known as conformational epitopes. The differences between these two epitope types lies in whether the epitope consists of contiguous amino acid residues in the primary structure or not, figure 6. Two recent studies on three-dimensional structures of more than one hundred antibody-antigen complexes showed that none of the studied epitopes were continuous, but rather consisted of multiple sections of continuous residues brought together in a three-dimensional space, forming discontinuous epitopes [212,220]. Also, both studies reported a similar number of residues to be part of en epitope, 18 to 19 residues [220] and an average length of 15 residues with a standard deviation of 4 residues [212]. The average maximum stretch of continuous residues in these discontinuous epitopes were reported to of approximately 5 amino acids, and 60% of all epitope residues were found to form at least 3 residues in a continuous stretches [212]. A common misconception is to call any short peptide, of e.g. 10 amino

acids, for a continuous epitope just because of its length. Even a short stretch of amino acids can be enough to form discontinuous epitopes [221,222].

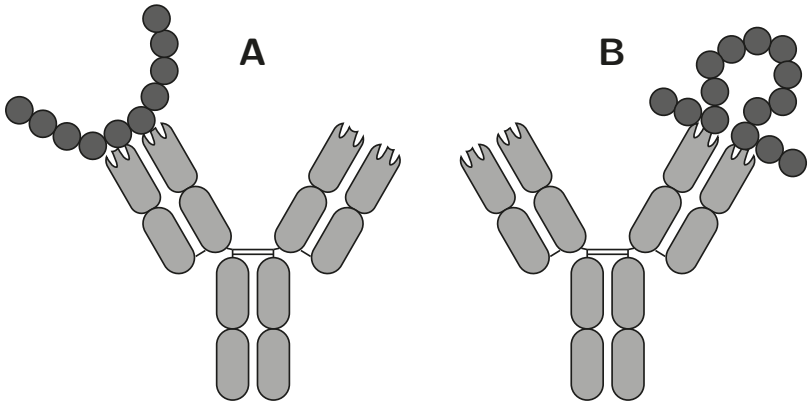


***Figure 6**. Antibody binding to a continuous epitope, where the amino acids are contiguous in the primary structure (A), and to a discontinuous epitope, where the amino acids are not contiguous in the primary structure (B).*

How about the paratopes? Antibody paratopes are, as described previously, constructed of CDRs located in six different loops, and thus all antibody paratopes are discontinuous by nature [221]. However, just as continuous stretches of epitopes exist, continuous stretches of paratopes have shown to possess similar properties for binding epitopes [223]. Also in studies by Laune *et al.*, the authors could show that 12-mer peptides covering contiguous regions of CDRs in the heavy and light chains of different antibodies, were capable of binding the different antigens with close to retained affinity [224,225]. Furthermore, the paratopes are, not surprisingly, composed of the same number of amino acid residues as the epitopes, on average 18 amino acids. Given that the total number of amino acid residues in the 6 CDRs of an antibody is 56 on average, only 1/3 of these are taking part in binding a respective epitope [220]. This means that an antibody has many other potential paratopes, which can consist of other amino acids in the CDR, but also have paratopes with overlapping sequences. As discussed in the previous chapter, an antibody can possess multiple specificities since different bindings are possible between antibody and antigen(s). Each paratope-epitope interaction can be divided into three different groups [201], 1) antibody is binding to an epitope of the antigen it was made against, 2) antibody binds to an identical epitope but on another antigen than the one it was made against and 3) antibody cross-reacts by recognizing an epitope that is structurally related (with no necessarily sequence similarity) to the epitope on the antigen it was made against. Given that there

are multiple paratopes for each antibody, each antibody may bind related and unrelated epitopes with limited sequence similarity, but also epitopes with related or unrelated structures, as opposed to the epitope on the antigen it was raised against [199,201,226]. This is called polyspecificity and is usually referred to as antibody promiscuity, as exemplified in a study by Keitel *et al.* [227] where X-ray crystal structures of an antibody Fab fragment with four different peptides were analyzed. They studied a peptide epitope from the antigen it was raised against, one homologous peptide, two unrelated peptides and observed that all the peptides bound to the Fab CDR loops however with different paratope interactions.

One particular class of epitopes that exists due to the polyspecific nature of proteins, including antibodies, is mimotopes. These mimotopes are defined as epitopes that differ in the primary structure from other epitopes, but essentially interact with the same paratope on the antibody [221,228,229]. Furthermore, mimotopes have been of great interest as they play an important role in autoimmune diseases [230,231]. The hypothesized mechanism is that antibodies or activated T-cells that are created during an immune response towards a pathogen, e.g. a virus, also are cross-reactive towards self-antigens. In turn they elicit a response where the immune system attacks its own host. If a self-antigen shares its mimotope with non-self-antigen, the self-antigen can break tolerance and thereby cause an autoimmune reaction, which eventually can cause disease [232]. Such events are based on cross-reactivity between antigens with the same mimotopes and are also called molecular mimicry [229].

Autoreactivity in autoimmune diseases

The immune system's remarkable ability to distinguish between what is pathogenic or not, is what ultimately starts an immune response, which in turn leads to the production of antibodies. As discussed in the previous chapters, this has been used to study biology for more than a century. However, in some cases, instead of protecting its host against pathogens, antibodies or even the whole immune system may be mistaken and recognize the host as a foreign threat, thereby starting an autoimmune response. Therefore, one of the greatest challenges for our immune system is to distinguish between self and non-self, in order to avoid an immune response towards self and thereby cause damage. There are over 80 different diseases suggested to have an autoimmune component, such as systemic lupus erythematosus, multiple sclerosis,

narcolepsy, and schizophrenia, with a total prevalence of 3-9% in the world population [149,233]. Depending on the etiology and underlying mechanisms of these autoimmune diseases, some have more, less or similar involvement of either of the B- and T-cells [234-236].

So what causes autoimmunity? The central processes in which an organism governs and protects itself against autoimmune reactions are referred to as immunological tolerance, and if there is a loss in tolerance, autoimmunity can be developed. As lymphocytes mature there are distinct mechanisms, known as the central tolerance, in which tolerance is induced in the B-cells in the bone marrow. In the thymus it induces tolerance in the T-cells. As previously discussed in chapter 3, one of the main mechanisms in creating the diverse B-cell repertoire is through somatic recombination of the immunoglobulin genes during B-cell development. However, due to its randomness in creating the diversity, it can also produce autoreactive B-cell receptors. Therefore during the B-cell development in the bone marrow when immature B-cells have undergone somatic recombination of their immunoglobulin genes, they have three alternatives: the first is to undergo a positive selection where the B-cell is allowed out to the periphery. Second, if the signal for positive selection is too low it can redo its light chain rearrangements. Third, if the B-cell receptor binds an antigen in the bone marrow it will undergo negative selection and go through apoptosis (although it can be rescued in some cases by a receptor editing mechanism).

For the T-cells, T-cell precursor cells from the bone marrow enter the thymus to slowly proliferate over a few weeks into thymocytes. Then the double positive thymocytes undergo positive selection by interacting with its T-cell receptor to cortical thymic epithelial cells that have highly expressed self-MHC class I and II molecules. Cells with high affinity towards self-MHC molecules result in negative selection and deletion, while no interaction lead to a lack of survival signals and apoptosis as a consequence. A very small fraction of cells with low to intermediate interaction survives and matures into single positive T-cells. Next, cells migrate to the medulla of the thymus and undergo negative selection by interacting with medullary thymic epithelial cells that, through the transcription factor AIRE, expresses a wide variety of proteins. These are then processed and presented on the MHC class I and II molecules. Once again, cells with high affinity to the MHC molecules are deleted except a subset that develop into regulatory T-cells, and remaining T-cells that do not bind the

MHC molecules survive, and exit the thymus and enter circulation. [149,237,238]

Neither of the above-mentioned mechanisms of central tolerance are perfect as some lymphocytes might not have been exposed to all self-antigens during selection. Additionally they might have been interacting with low affinity and thereby managed to escape cell death. Therefore, autoreactive lymphocytes can be present outside the thymus and bone marrow. Fortunately, mechanisms in the periphery, peripheral tolerance, can regulate these autoreactive cells. First, as discussed in the previous chapters, during B-cell activation in the secondary lymphoid tissues, the B-cells require activation through T-cells in order to activate into antibody-producing cells. Therefore, the B-cells will not be able to become activated, and will thereby become anergic. In a similar way, the T-cells require additional activation by antigen presenting cells aside from the MHC and T-cell receptor interaction; e.g. through a co-stimulatory signal to CD28 on the surface of the T-cell as well as cytokine signaling, and if any signal is absent the T-cell will become anergic. In addition, the regulatory T-cells can regulate other T-cells in the periphery by different mechanisms, such as deprivation of local cytokines. Thereby they shut off any T-cell proliferation in its vicinity, secrete inhibitory cytokines, inhibit antigen-presenting cells that presents an autoantigen, or directly kill cells in its surrounding.

If the above-mentioned mechanisms fail to induce tolerance, the self-reactive lymphocytes can then cause autoimmunity. As summarized in a review by Theofilopoulos *et al.* [239], autoimmune diseases are multifactorial and rely on a variety of factors for the development of the disease, including genetic (and epigenetic) factors and environmental factors. The genetic factors that are associated with increased susceptibility of getting autoimmune diseases include, certain HLA alleles such as HLA-DQB1*0602 which is associated with narcolepsy, type 1 diabetes [240], but also non-HLA genes such as in systemic lupus erythematosus that is associated to a certain haplotype of the transcription factor, interferon regulator factor 5 [241]. Furthermore, a recent review by Cho & Gregersen [242] found that more than 200 genetic loci have been associated with one or more autoimmune diseases, including cytokines, cytokine receptors, genes involved in innate immunity, transcription factors and many more. To further add complexity to genetic impact, concordance rates in twins for acquiring several autoimmune diseases have been studied. It was been shown that there is 12-33% concordance in identical twins, compared to 0-4% in non-identical twins [243]. Although this clearly shows the genetic

contribution to autoimmunity, it does not explain all the cases, and thus the divergence have been suggested to be due to environmental factors [243]. Environmental factors that are suggested to be involved in the pathogenesis of autoimmune diseases include: cigarette smoking which increases the risk of getting rheumatoid arthritis [244,245] and systemic lupus erythematosus [246], deficiencies of vitamin D, e.g. through diet, have been associated to increased incidence of autoimmune diseases such as in multiple sclerosis [247]. Furthermore, 78% of all patients that have autoimmune diseases are women [248]. This so-called gender bias has been attributed to hormones that are usually found in higher levels in women, estrogen and prolactin, which cause e.g. survival of autoreactive B-cells in SLE [249]. Also, it has been shown that dysfunctional composition of the gut microbiota, due to sex hormone levels, can cause a variety of autoimmune diseases [250], including inflammatory bowel disease [251]. In addition to the HLA gene associations in narcolepsy, it has been shown that influenza type A H1N1 and particularly certain vaccine preparations of that strain (Pandemrix ASO3-adjuvant) can cause narcolepsy in genetically susceptible populations [252].

As discussed, there are a variety of different mechanisms in which the human body protects itself against autoreactivity, and there are equally varied genetic and environmental factors that increase the susceptibility for autoimmunity. However, the triggers of autoimmune diseases have been attributed to dysfunctional gut microbiota and various infections, with inflammation as a common denominator [239,251,253]. As discussed earlier in this chapter, molecular mimicry is one of the mechanisms that can induce autoimmunity, but there are also three additional mechanisms including intermolecular epitope spreading, bystander activation and release of cryptic antigens, and these can affect both B- and T-cells [254-256], since B-cells rely on T-cell activation. A suggested mechanism for molecular mimicry is that a pathogen first causes a cross-reactive T-helper cell to be activated by an antigen presenting cell, which promotes inflammation through cytokine and chemokine release and thus results in tissue damage [255]. The inflamed tissue can then release proteins (and other molecules) that are mimotopes with the pathogen, and thereby start an autoimmune response. One example of this is multiple sclerosis where Theiler's encephalomyelitis virus has been suggested to mimic myelin proteolipid protein and to cause experimental autoimmune encephalomyelitis (EAE) in mice [257]. In bystander activation, self-reactive T-cells become activated by an ongoing tissue infection by e.g. a virus similar to that in molecular mimicry, which leads to inflammation and recruitment of self-

reactive T-cells that can start an autoimmune response due to presentation of certain self-antigens [255]. This has been suggested to cause type 1 diabetes by Coxsackie virus infection [258]. In contrast to molecular mimicry, bystander activation is independent on what the initial infectious agent is as it relies on self-reactive T-cells rather than cross-reactive [255]. The third mechanism is the release of cryptic antigens, that is release of self-epitopes that have previously not been seen the by the immune system [255]. During an infection, pathogen-specific T-cells can induce changes in antigen presenting cells. Just as when antigen-presenting cells engulf infected cells, new types of the self-peptides can be presented and thereby trigger an autoimmune response [231,255]. Presentation of cryptic antigens have been suggested to play a significant role in certain autoimmune diseases, such as in EAE [231,259,260]. The fourth and last suggested mechanism on how infections can cause autoimmunity can be attributed to a process called epitope spreading. Here a chronic infection due to a pathogen can cause a diversification of T-cells by releasing self- and pathogen-specific peptides that are presented by antigen-presenting cells, which can activate self-reactive T-cells [255,261]. Epitope spreading has been suggested to be involved in various different autoimmune diseases, including type 1 diabetes, multiple sclerosis, rheumatoid arthritis and systemic lupus erythematosus [261,262]. However, it should be emphasized that these mechanisms interplay and that a combination of these mechanisms can induce autoimmunity [255].

Sixty years ago, in 1957, the first defining criteria for autoimmunity, Witebsky's postulates, were described [263]. Since then, there has been a plethora of identified autoantigens for B-cells [264] and T-cells [255,265]. Detection of these autoantigens can be used for diagnosis and therefore better treatment of patients, for example detection of anti-nuclear autoantibodies in systemic lupus erythematosus [266] and detection of autoantibodies in breast cancer [267]. Identification of autoantigens can also help in the development of new treatments, e.g. through immunomodulation [268,269] and performing autoantigen-specific B-cell depletion [270]. Autoantigens can also be used to monitor disease progression over time by for instance detecting anti-citrullinated protein antibodies in rheumatoid arthritis [271] or detecting intra-molecular epitope spreading of certain autoantigens as a proxy for relapse and remission of disease, as in an animal model of multiple sclerosis [272]. Furthermore, there are a variety of methods and strategies for identifying and characterizing T- and B-cell autoantigens and ultimately their epitopes, which will be described in the following section of this chapter.

Epitope mapping and identification of autoantigens

B-and T-cells autoantigens and epitopes are important to study in the context of autoimmunity. Antibody epitopes are also valuable to determine to further characterize interaction partners of an antibody and thereby its selectivity. Strategies for identifying and determining such interactions can be divided in three categories, i.e. function-based methods, prediction-based methods, and structure-based methods, although they can also be used in conjunction with each other.

As we have seen in the first section of this chapter, three-dimensional antibody-antigen structures can give information about the antibody paratope and epitope, and similarly it is possible to study the T-cell epitopes (MHC molecules in complex with peptides) [273]. This is usually done by X-ray crystallography, nuclear magnetic resonance or electron microscopy, and the resulting structure of the complexes are analyzed. Atoms from the antibody and antigen that are separated by less than 4 Å are usually regarded as being part of an epitope or paratope [274]. Moreover, structural epitopes provide a very detailed picture, with atomic resolution of the epitope-paratope interactions, and is therefore a very valuable tool. However it is also considered to be a laborious and expensive method for mapping epitopes [275]. In addition, for some antigens, such as membrane proteins it has been proven to be particularly difficult to generate such structures [276].

Another way to experimentally map epitopes involves mapping so-called functional epitopes, which are defined are based on whether the antibody exhibits any binding to them [277]. These epitopes are usually shorter than the structural epitopes, since not every amino acid in the structural epitope is needed to bind in order to provide enough binding energy between the antibody and the antigen [222,278]. As shown by Dougan *et al.*, a few amino acids can constitute more than half of the free energy of a paratope-epitope binding in an antibody [279]. There are many strategies for mapping the functional epitopes and in the following section these different strategies for determining and predicting epitopes will be described including methods for identifying autoantigens, with focus on array-based technologies.

Site-directed mutagenesis

An approach to map the functional epitope is to perform site-directed mutagenesis of the target protein. Here, knowledge of the protein sequence is needed in advance in order to map the epitopes, as the protein variants need to be cloned and produced. One can either systematically mutate all amino acids of the antigen, one amino acid at a time, or preferably make an educated deduction of which sites to mutate and produce only a handful of variants that are of interest [277,280]. Amino acids are usually mutated (or substituted) to alanine residues [277] since it is regarded to be inert for the conformation of the protein or peptide as it is present in a variety of secondary structures [281]. By analysis with ELISA or SPR, changes in the binding of the antibody to the mutated antigen can be compared to the native antigen and thereby pinpoint which amino acids that are essential to the binding of the antibody and ultimately determine the epitope [280,282].

Mass spectrometry-based methods

There are two fundamentally different strategies in order to determine epitopes using MS [283,284]. The first type of methods relies on using proteases to reveal the epitopes, either by excision or extraction [283,284]. In so-called epitope excision, the antibody is immobilized on a solid surface and then incubated with the antigen and subsequently with a protease, e.g. lysyl endopeptidase or trypsin [284,285]. Depending on where the antibody has bound the antigen, the digestion efficiency will be lower due to steric hindrance, and thus expose the epitope in complex with the antibody [283,284]. By washing away unbound peptides, the epitope sequence can be analyzed by e.g. MALDI-MS [283,284]. In epitope extraction, the antigen of interest is first digested by proteases before incubation with the antibody [283,284]. Similarly to epitope excision, by washing away unbound peptides, the epitope can be eluted and be analyzed by e.g. MALDI-MS [283,284]. The other type of methods relies on chemical modification, either of the surface-accessible amino acids of the antigen and antibody, or of the amino acids in the interacting surfaces, i.e. the epitope and paratope [283,284]. One example of these methods is hydrogen/deuterium exchange [286] in which the antigen and antibody are first incubated individually in deuterium oxide. This causes an exchange of the hydrogen atoms to deuterium in the amides of the solvent-exposed parts of the protein backbone [286]. After incubating the antigen and antibody together, the solvent is changed back to water and which reverses the incorporation of deuterium in the surface exposed residues while the amino acid residues in the epitope and

paratope retain their deuterium. After quenching the exchange reaction in low pH, proteolytic cleavage and subsequent analysis using e.g. ESI-MS will detect and quantify the mass-shift introduced by the two different isotopologues, and from this the epitope can be deduced [286].

Array-based methods

Robert Bruce Merrifield introduced the solid-phase peptide synthesis (SPPS) in 1963, which enabled covalent coupling of amino acids on solid surfaces. The technique has since then improved greatly and has in the last decades constituted the foundation of array-based epitope mapping [287]. By synthesizing multiple peptides, it is possible to create a library of peptides and there are a variety of approaches in epitope mapping of antibodies using these libraries [288]. A common approach introduced by Atassi *et al.*, [289] is to synthesize a peptide library representing the antigen of interest, usually using overlapping peptides, and then measure the binding of the antibody to each peptide separately. Another commonly used method, introduced by Geysen *et al.* [290], relies on synthesis of peptides representing the antigen, but also to perform complete substitution of every amino acid in all peptides, thus creating 20*n different peptides for a n-mer peptide. Although a complete substitution is a rather exhaustive method, it enables identification of the essential amino acids for the epitope of the antibody. Less thorough methods compared to complete amino acid substitution exist, such as alanine substitution analysis [281], were each amino acid in each peptide is substituted to alanine, thus creating n+1 peptides for an n-mer peptide. In a similar way to the previous method, by analyzing how the amino acid substitution affects the binding of the antibody, it can be used to determine the epitope. Furthermore, it is also possible to do truncations of the peptides in the N-terminus or C-terminus in order to create a peptide library of different length, and thereby enable determination of the minimal epitope length [288]. These types of experiments have traditionally been carried out either using ELISA. The work by Geysen *et al.* in 1984 [290] is often regarded as the first attempt of creating array analysis of epitopes and therefore also the forerunner to the current peptide array technologies [291].

Peptide array technologies started to appear in the 90's, which enabled epitope mapping in highly multiplexed assays [291,292]. Furthermore, developments of photolithographic techniques that were first introduced in 1991 by Fodor *et al.* [293], have now enabled synthesis of millions of peptides. These report up to 10

million peptides per array, immobilized on solid surfaces equal to the size of a regular microscope slide [294,295]. Therefore, such arrays can be used for large-scale analysis of epitopes and for instance enable analysis of the antibody repertoire in infectious diseases [296] or map the immune response in cancer [297].

As these above-mentioned approaches typically uses peptides that are built on 12-20 amino acids, an obvious drawback in using these is that discontinuous epitopes might go undetected since the average span of epitopes in the primary structure are reported to be between 50-79 amino acids [220]. Already in 1993 Luzzago *et al.* [298] showed that it is possible to construct shorter peptides that mimic discontinuous epitopes. Since then researchers have used random collections of peptide sequences to identify mimotopes, e.g. for antibodies towards hepatitis C virus in human sera [299], but also mapping other protein-protein interactions [300]. However, these techniques used to rely on using surface display techniques, and specifically bacteriophage display since these methods require large libraries of peptide sequence [300], see next section of this chapter. However, in a recent article, these methods have also been applied to peptide microarrays. By covering 83% of all possible tetramer and 27% of all possible pentamers, the authors used the peptide microarray to create so-called immunosignatures that can distinguish antibody responses from patients with various infectious diseases with the same array design [301]. Therefore, one can argue that although peptides themselves cannot constitute the same three-dimensional structure as the native antigen of an antibody, they can still be able to create mimotopes that can detect them. Also, as discussed previously in this chapter, each discontinuous epitope consist of several continuous epitope parts with average maximum stretch of approximately 5 amino acids [212], therefore these peptide array designs offer great opportunities for epitope mapping of antibodies.

As discussed in the previous section, identification of autoantigens and thereby potential autoantibodies, can serve as valuable markers to monitor disease progression or for diagnosis. Given the heterogeneity of many autoimmune diseases and the lack of biomarkers, it is of great interest to profile the autoantibodies and autoantigens to further increase the understanding of such diseases [302]. In addition to the peptide arrays, another way of profiling autoantibodies is by using proteins, either purified full-length proteins or protein fragments that are immobilized on a solid surface. Unlike using peptide arrays, a bottleneck in using protein arrays has been the availability of

reagents, and specifically to produce proteins in high-throughput, in terms of cloning, expression and purification of the proteins. These efforts have previously been hampered by the lack of full genome sequence [303,304]. Since the introduction of so-called proteome-wide protein arrays almost two decades ago [303,304], there have been several large-scale initiatives for production of proteins including development of protein production techniques [305,306]. In addition, several approaches have been developed to circumvent the tedious tasks of cloning and purification. These rely on proteins on the array being translated in a cell-free environment *in situ*, such as DNA array to protein array [307] and nucleic acid programmable protein array [308]. Nevertheless, there are still no protein arrays with full coverage of the human proteome yet, although full-proteome coverage is indeed on the horizon [304,309-311]. Therefore one can argue that peptide arrays might be a more complete tool in analyzing autoantibodies towards the whole proteome. However, and despite non-proteome coverage of protein arrays, the technology has been used to make many interesting discoveries of potential autoantigens in different diseases that could be of clinical use, thus showing that it is a powerful tool [311-315].

Surface display-based methods

As mentioned earlier, researchers already used different phage display methods in 1993 for epitope mapping. In fact, already in 1985, the properties of filamentous phage were discovered, and enabled the display of cloned antigens on the surface of the phages [316]. Since then, many different display techniques have been developed including, ribosome display, mRNA display, yeast cell display, as well as bacterial display [317,318]. Due to both the flexibility of the preparation of the libraries regarding length of displayed molecule, and the typical maximum size of the libraries, it is nowadays possible to create a library that enables analysis of complete proteomes represented in peptides and protein fragments. For instance, in a study by Larman *et al.* in 2014 [319], they designed a library for phage display, consisting of 413,611 unique peptides with the length of 36 amino acids, covering the whole human proteome. They used the library to analyze autoantibodies, first by incubating the phages with the patient samples, then immunoprecipitating the antibodies followed by PCR and sequencing as read-out of the phage clones. Thereby, they could epitope map autoantibodies from three patients on 36-mer peptides representing the whole human proteome.In another recent study [320], Elledge *et al.* used a similar approach to map antibody response towards the whole human virome (206 different species). They created a phage library that

displayed 93,904 unique 56-mer protein fragments with 28 amino acid overlap in 569 human samples. Furthermore, their study showed that they could use their phage display technique to explore and study immune responses towards the virome in a high-throughput manner [320]. It is now possible to use surface display-based methods to analyze proteomes in highly multiplexed and automated fashion through developments of bacterial display [318] as well as the techniques described above by Larman *et al.* and Elledge *et al.* [319,320].

Prediction-based methods for epitope mapping antibodies

The first protein structure was resolved in the 50s [321], and in 1974 Chou and Fasman [322] could take advantage of 15 protein structures that were resolved in order to predict protein conformation. Inspired by these findings, Hopp and Woods presented a method to predict antibody epitopes in 1981 [323]. It had been shown that epitopes usually are found on residues that are exposed to water. So they used previously tabulated values for hydrophilicity of each amino acid and then calculated moving averages of hydrophilicity in six residue-windows along the protein sequences [323]. This was then applied to 12 proteins, and the local maximums in the hydrophilicity scores were predicted to be continuous epitopes and by that they could conclude that they successfully predicted the epitopes [323].

Since then, several attempts have been made to create better prediction models for prediction of continuous epitopes by incorporating different physio-chemical properties of the amino acids, including hydrophilicity [324], flexibility [325], solvent accessibility [326] but also antigenicity [327] as discussed previously [328]. A systematic review made in 2004 of 484 different propensity scales for prediction of continuous B-cell epitopes could not find that any of the 10 million combinations they tested contributed to any significantly higher prediction than random [329]. In the following years, more sophisticated B-cell epitope prediction methods were developed that used propensity scales and machine learning methods, such as in the prediction tools [328]: ABCpred, [330] BCPred [331] and BepiPred 1.0 [332]. As discussed by Fleri *et al.* [333], BepiPred 1.0 is the best performing prediction method for predicting B-cell epitopes, and although it performs significantly better than random it is still limited in the predictive power of epitopes [328,332]. Recently, a new version of the BepiPred (2-0) was released, based on a Random Forest algorithm that is solely trained on data from three-dimensional structures of epitopes, which

improves the classifications of epitopes and thus outperforms the former version, however there are still room for increased predictive power [334].

Furthermore, as the availability of three-dimensional structures of antibodies and antigens have increased over the last decades, more sophisticated tools for predicting discontinuous epitopes have also developed [328]. However, in contrast to the prediction of continuous epitopes, when predicting discontinuous epitopes of a protein, the three-dimensional structure of the antigen (or similar) must be determined in order to enable such predictions [335-337].

In silico-based methods for predicting antibody epitopes can be very valuable tools, and especially in conjunction with experimental work. As more three-dimensional structures of antibodies and antigens will be determined; accordingly the predictive power and conclusions of prediction-based methods will grow enormously.

Present investigations

The work presented in this thesis aims to identify and characterize potential disease-relevant antibodies, IgGs, in the suggested autoimmune diseases, narcolepsy and psychosis (**Paper I** and **Paper II**), in the autoimmune disorder multiple sclerosis (**Paper III**) as well as to characterize antibodies used to study the *Plasmodium falciparum* parasite (**Paper IV**). This work has been performed thanks to the recent advances of high quality tools for affinity proteomics, such as the high-density peptide arrays (**Paper III** and **Paper IV**), and the massive resources of recombinant proteins (and antibodies) from the Human Protein Atlas Project (**Paper I** and **Paper II**).

A common theme for **Paper I - Paper III** is a three step process consisting of i) broad untargeted screening of reactivity in a small set or subset of patient samples, using arrays of in total thousands of protein fragments or peptides, ii) selection of top reactive proteins or peptides for the creation of a targeted array, which then is used to profile a larger set (not for **Paper II**) of clinical samples, and iii) detailed investigation of the amino acid composition of those antigens or peptides that display a disease related differential autoantibody reactivity.

In **Paper I** and **Paper II**, human protein fragment arrays from the Human Protein Atlas workflow are used, whereas whole proteome peptide arrays were used as a starting point in **Paper III**. **Paper IV** continues the work on high-density peptide arrays, by exploring reactivity towards pathogenic antigen peptides. Since the peptide arrays' features (spots) are so densely packed on the surface, a special workflow was established for the analyses of those arrays, both for the image analysis but also subsequent data analysis.

Paper I: Autoantibody targets in vaccine-associated narcolepsy

Narcolepsy with cataplexy is a chronic disease characterized by low levels of the neuropeptide hypocretin (also known as orexin) in the cerebrospinal fluid of affected individuals [240]. This has been attributed to hypocretin-producing cells in the hypothalamus, which is responsible for the sleep-wake cycles. However, the exact cause for the loss of these cells is not known, although mechanisms suggesting an autoimmune reaction are getting increasingly support from the scientific community [338]. One of the reasons is that there is a strong HLA association to narcolepsy, in which the allele HLA-DQB1*0602 is found in 98% of all narcolepsy patients [339]. Also, shortly after the spread of the pandemic H1N1 influenza followed by the mass vaccination campaigns, higher incidences of narcolepsy were reported in several countries, including Sweden and Finland [340]. Therefore there has been suggestion of molecular mimicry as a potential pathway of the disease development [338].

Due to the proposed mechanisms of narcolepsy, we profiled the antibody repertoire of individuals with vaccine-associated narcolepsy, sporadic narcolepsy, as well as controls, in order to reveal potential disease-relevant autoantibodies. Initially we investigated antibodies in serum samples from a Finnish sample collection consisting of 39 narcolepsy patients, whereof 20 vaccine-associated, 16 sporadic and three without knowledge of association, as well as 18 controls. We profiled the repertoire of autoantibodies by using planar antigen arrays composed of in total 10,846 unique PrESTs, representing 7953 human proteins. By studying the general antibody reactivity patterns of the analyzed samples on the PrESTs, a large individual heterogeneity could be observed, where a majority of the reactive antigens had only single individuals that were reactive towards them, while a few antigens were reactive in more than half of the analyzed individuals (Figure 7).

Based on antibody reactivities, 244 PrESTs were selected for differential reactivity frequencies when comparing the narcolepsy and control groups, as well as between vaccine-associated and sporadic narcolepsy. A targeted bead array was created composed of these antigens, and the same samples were analysed on the bead array platform as a technical verification, resulting in 14 antigens with verified differential reactivity between the patient groups. These were further validated in a second stage, on an independent Swedish sample collection consisting of plasma samples from 117 controls, 59 narcolepsy

patients, whereof 46 that had vaccine-associated narcolepsy, 4 with sporadic narcolepsy and 9 without knowledge of associations. Two of the 14 antigens were found to have higher reactivity in the narcolepsy patient group in both sample sets, namely cytosolic 5-nucleotidase 1A (NT5C1A) and methyltransferase-like protein 22 (METTL22), Figure 8. One of the reasons for the discrepancy between the first stage of discovery and the second stage of verification could be the lack of sample material from individuals with sporadic narcolepsy in the second sample collection.

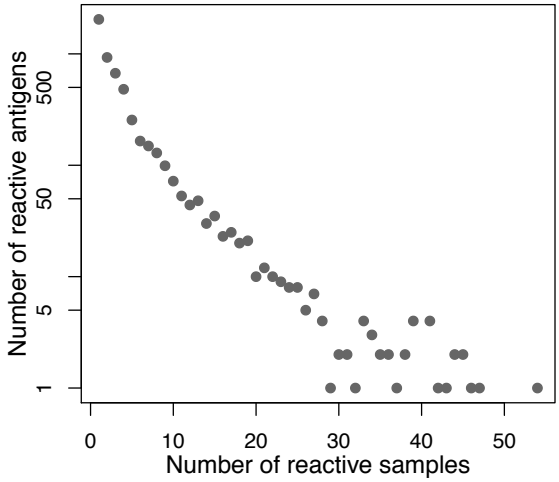


Figure 7. *Antibody reactivity overview of the 57 patient samples on 10,846 human protein fragments, with the majority of reactive antigens were found in single individuals.*

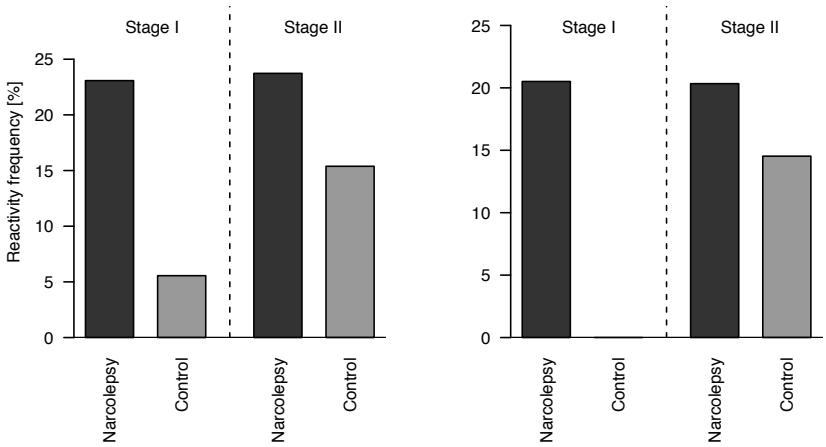


Figure 8*. Percentages of patients with antibody reactivity towards METTL22 (left) and NT5C1A (right). Antibody reactivity towards the two proteins was found to be higher in the narcolepsy patients compared to controls in two independent narcolepsy cohorts.*

In order to investigate the potential molecular mimicry, antigens with higher reactivity in the vaccine-associated narcolepsy group were divided *in silico* to

12-mer peptides and searched with BLAST against the influenza H1N1 proteome, with the same strain found in the Pandemrix vaccine (A/California/7/2009). This resulted in a few stretches of sequence of high sequence identity when comparing the human antigen sequence and the H1N1 sequences, e.g. the antigen representing glutaminase 2 (GLS2) had a stretch of sequence with high sequence identity to the nucleocapsid protein of H1N1.

In summary, in this study the antibody repertoire in patients with narcolepsy were profiled on human protein fragment arrays. In total 10,846 different PrESTs representing 7,953 protein-coding genes were used to profile potential autoantibodies. After technical and biological verification, two antigens, NT5C1A and METTL22, were identified as potential autoantigens due to higher IgG reactivities in the narcolepsy sample groups compared to controls. NT5C1A has previously been characterized as an autoantigen in sporadic inclusion body myositis [341]. In addition, a few antigens that were found to have higher IgG reactivities in the narcolepsy group did have short stretches of high sequence identity to the nucleocapsid protein of H1N1, suggesting potential cross-reactivity from H1N1 vaccine and/or infection.

Paper II: Untargeted screening for novel autoantibodies with prognostic value in first-episode psychosis

Ever since the first autoantibodies were observed in patients with schizophrenia in 1937 [342], there have been several reports discussing whether there are any immunological abnormalities in schizophrenia or not [343], including several reports on autoantibodies [344,345]. Therefore, in order to investigate this further, autoantibodies from patients with a first-episode psychosis were profiled on the in-house developed human protein fragments. Plasma samples were obtained from 53 patients with a first-episode psychosis and 41 non-psychotic controls. The disease progression, health status and psychiatric diagnoses, of the patients with first-episode psychosis were followed for in average seven years after sample collection.

Similar to **Paper I**, PrEST arrays were utilized, where in total 2304 PrESTs representing 1812 unique protein-coding genes were analyzed. Twenty-nine antigens showed significantly different antibody reactivity when comparing the first-episode psychosis patient group and non-psychotic controls. Similar to **Paper I**, there was a large individual heterogeneity, with a few antigens

reactive in almost all samples while most antigens were either not reactive or reactive in only single individuals, and in addition reactivities could be seen to similar extent in both groups, regardless of diagnosis.

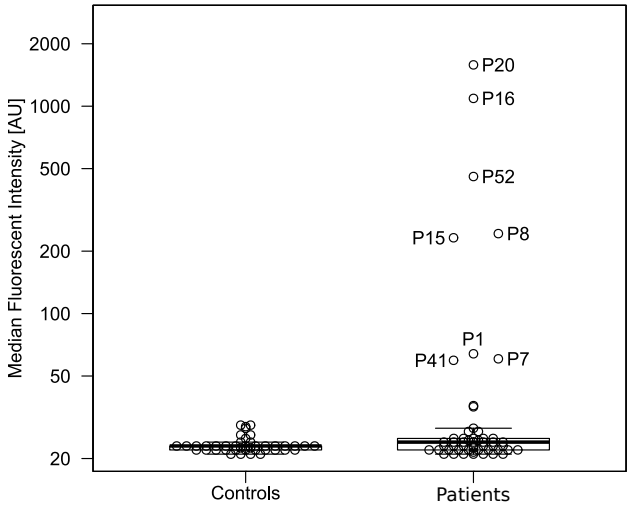


***Figure 9**. Antibody reactivity towards the N-terminal fragment of PAGE2B, as defined by a sample-specific cutoff, was found to be higher in patients that developed schizophrenia after the first-episode psychosis compared to controls.*

Furthermore, the 29 selected PrESTs were further analyzed on suspension bead arrays as a technical verification, together with 26 PrESTs covering different regions of the protein targets. Results from the technical verification revealed one antigen with a disease-associated reactivity. This antigen, representing three proteins with very high sequence identity, namely the P antigen (PAGE) protein family (PAGE2/PAGE2B/PAGE5) showed significantly higher antibody reactivity towards the antigen in the first-episode psychosis group (n = 8) compared to the non-psychotic group (n = 0), Figure 9. This protein family, which span between 111-130 amino acids, is fairly unknown in literature, with no publications of studies on protein levels. The identified antigen consisted of 39 amino acids and covered the N-terminal portion of the PAGE protein and in order to further explore the reactivities towards the N-terminal fragment of PAGE, a recombinant full-length protein of PAGE2B was used to perform an ELISA similarly to the analyses on the arrays. Apart from a few samples, the antibody profiles towards the full-length PAGE2B correlated well with the N-terminal portion of the PAGE proteins. In addition, the discrepant reactivities between the full-length PAGE2B and the N-terminal fragment of PAGE were through additional analyses attributed to reactivities towards a protein fragment representing the C-terminal portion of PAGE.

Based on the follow-up on health status and psychiatric diagnoses from the first-episode psychosis, some patients achieved complete remission while some developed schizophrenia, schizoaffective disorder, bipolar disorder, delusional disorder and unspecified nonorganic psychosis. In fact, out of the 14 patients that developed schizophrenia five of them showed antibody reactivity towards PAGE, while the remaining three PAGE-reactive patients that showed reactivity achieved complete remission, Figure 10. Thus, antibody reactivity towards the N-terminal portion of PAGE could be associated with higher risk of developing schizophrenia during the follow-up period.

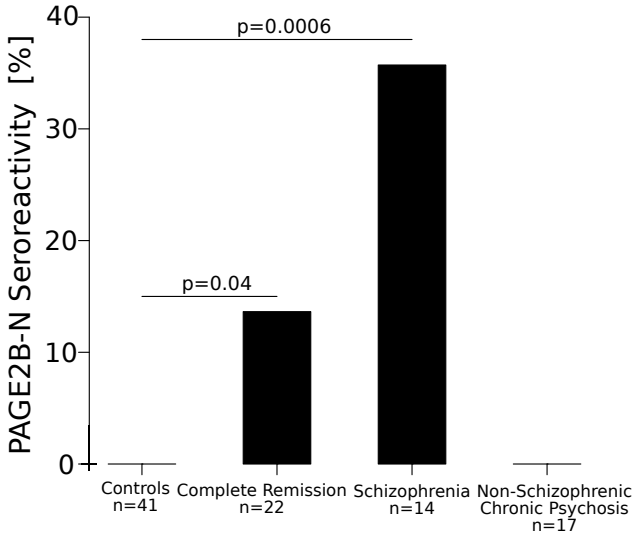


***Figure 10**. Higher reactivity towards the N-terminal fragment of PAGE2B was observed in patients that developed schizophrenia after the first-episode psychosis.*

In order to further characterize the reactivity, both immunohistochemistry analyses and epitope mapping of the eight PAGE-positive samples were performed. The immunohistochemistry analyses of PAGE in cerebral tissue showed positive staining by the affinity-purified polyclonal rabbit antibodies from HPA raised against the same N-terminal portion of PrEST as was used in the autoantibody analyses. The epitope mapping revealed a continuous stretch of six amino acids, NDQESS, which when BLAST searched for sequence similarity to other human proteins only revealed the three PAGE proteins with 100% identity.

In summary, in this study the antibody repertoire of patients with first-episode psychosis were profiled on antigen arrays consisting of 2,304 human protein fragments representing 1812 protein-coding genes. Then, an antigen representing the N-terminal region of three proteins with high sequence

identity, PAGE (PAGE2, PAGE2B and PAGE5), was identified to have higher IgG reactivities compared to the non-psychotic controls. Reactivities towards this antigen, was associated with higher risk of being diagnosed with schizophrenia following a first-episode psychosis. Further characterization of those IgG reactivities was made by studying IgG reactivities towards a C-terminal region of PAGE, as well as a representative full-length PAGE, and peptides representing a smaller region of the N-terminal fragment of PAGE. As a final step, immunohistochemistry analysis revealed that there might be expression of PAGE in the brain.

Paper III: Whole-proteome peptide microarrays for profiling autoantibody repertoires within multiple sclerosis and narcolepsy

As previously mentioned regarding **Paper I**, narcolepsy is a neurological disease with a suggested autoimmune component. Multiple sclerosis it is on the other hand well established to be an autoimmune disease, and it is the most common autoimmune disease involving the central nervous system [346]. Multiple sclerosis is characterized by an autoimmune attack on the myelin sheath, where the two major constituents, myelin basic protein and proteolipid protein, are targeted by the immune system, resulting in destruction of the myelin sheath and various symptoms affecting the central nervous system [346]. Similarly to narcolepsy, the self-destructive immune response is hypothesized to be due to molecular mimicry by several viruses including adenovirus and Epstein-Barr virus [347].

In this work, high-density peptide microarrays containing peptides representing the whole human proteome, as defined by the Consensus Coding Sequence project version 37.1 [348], were used for epitope mapping of the antibody repertoires in serum and plasma samples from patients diagnosed with multiple sclerosis and narcolepsy. The peptide microarrays consisted of in total 2.2 million 12-mer peptides with 6 amino acids overlap. Due to their high multiplexing capacity the sample throughput was initially low. Ten different plasma and serum samples from three controls, six narcolepsy patients and one sample pool consisting of ten secondary progressive multiple sclerosis (SPMS) patients, were analyzed on the peptide microarrays.

Based on this initial screening thousands of peptides were found reactive in each sample. But as discussed for **Paper I** and **Paper II**, most antibody reactivities were found in single individuals, and there were only a few antibody reactivities that overlapped, Figure 11. However, there were also peptides that had differential antibody reactivity between the narcolepsy, multiple sclerosis and controls samples. Those 14,082 peptides representing 1588 proteins, which showed differential reactivity, were further synthesized on a targeted design peptide microarray. These peptides, supplemented by peptides found reactive in at least 80% of the samples, peptides from proteins reported to be potential autoimmune targets within multiple sclerosis, and peptides derived from previous in-house autoimmunity profiling were synthesized on the new array of 174,000 12-mer peptides with 11 amino acid overlap. In total 23 plasma and serum samples were analyzed on these arrays. The sample set consisted here of eight patients with clinically isolated syndrome (a distinct disease similar to multiple sclerosis), eight patients with secondary progressive multiple sclerosis and seven patients with narcolepsy.

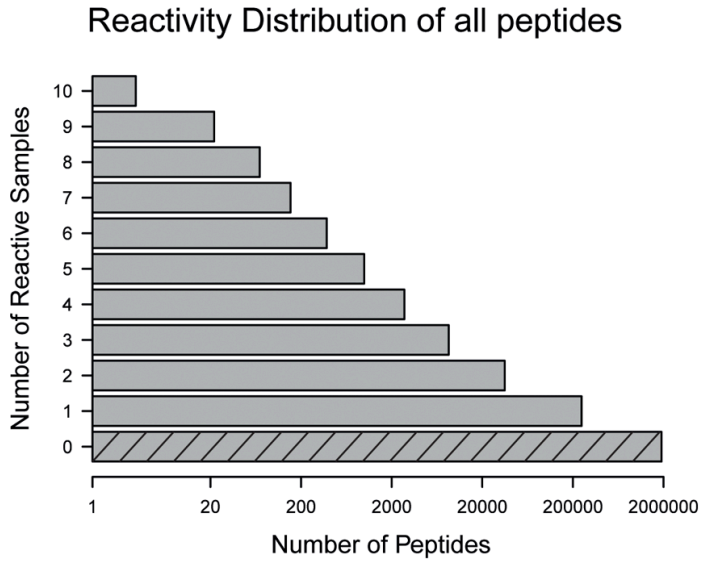


***Figure 11**. Large individual heterogeneity in antibody reactivity towards the 2.2 million peptides among the analyzed patient samples.*

At this stage, several peptides that were found reactive in the majority of the 23 analyzed samples, contained mainly the amino acids glycine and alanine, likely to be due to Epstein-Barr virus infection, and specifically reactivities towards the Epstein-Barr nuclear antigen 1 (EBNA1), which contain many glycine-alanine repeats. However, 133 peptides were identified with differential reactivity between the patient groups. These were used to create a peptide bead array, which enabled analysis of in total 448 plasma and serum samples.

One of these peptides represented Neurexin-1-alpha, a brain-enriched protein, which had significantly higher reactivities in the narcolepsy samples compared to the control samples, Figure 12. In addition, antibodies towards a peptide representing mitogen-activated protein kinase kinase kinase 7 (MAP3K7), a protein involved in inflammation, was found to be significantly higher in samples from relapsing remitting multiple sclerosis, secondary progressive multiple sclerosis, and in clinically isolated syndrome, compared to the reactivities in the samples from the control group (other inflammatory neurological diseases), Figure 12.

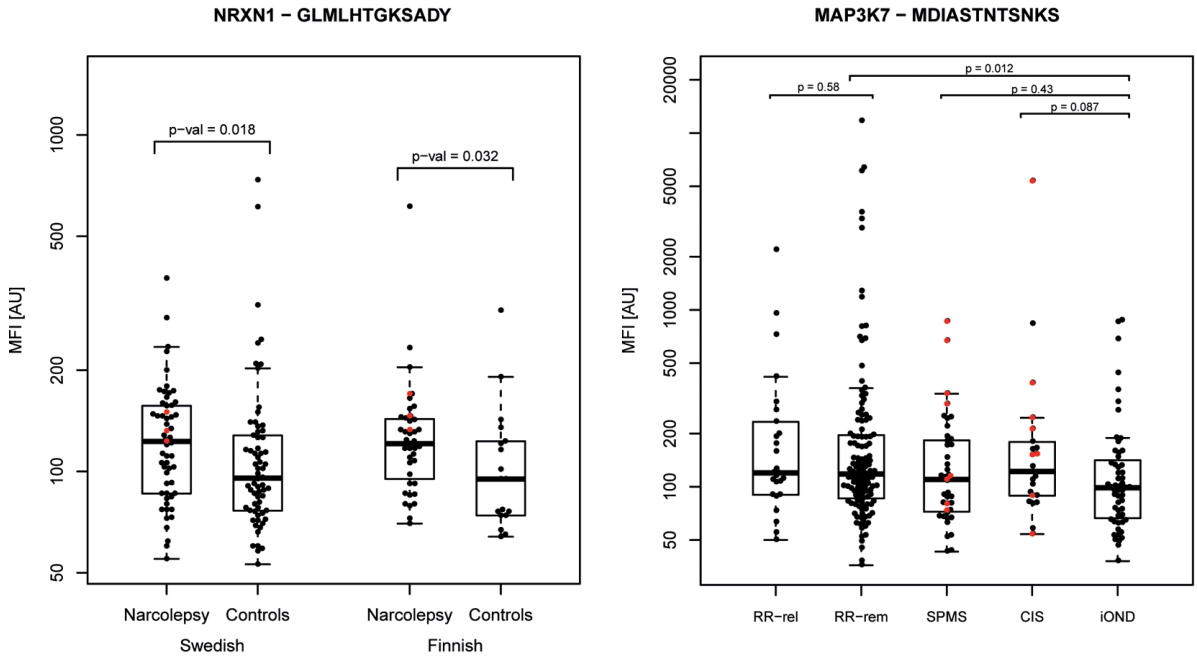


Figure 12*. A peptide representing NRXN1 had significantly higher antibody reactivity towards it in narcolepsy samples compared to controls in two independent sample collections (left). In two different multiple sclerosis disease groups, including CIS, had higher antibody reactivity towards a peptide representing MAP3K7 (right).*

In this study, antibody repertoires were profiled on three different arrays in patient samples from two autoimmune diseases, narcolepsy and multiple sclerosis. Initially whole-proteome microarrays with 2.2 million 12-mer peptides with 6 amino acid overlap were used for the profiling. Peptides giving differential IgG reactivity were selected for further analysis on an array design with less peptides (174,000) per subarray, but that enabled more samples to be analyzed. From these analyses, peptides giving IgG reactivity found in all patients were glycine and alanine rich, suggesting cross-reactivity to EBNA1. A few peptides that showed differential IgG reactivity were selected to be analyzed on a bead-based peptide microarray, enabling hundreds of samples to be analyzed in one assay. The resulting experiments on bead-based microarray

revealed two potential autoantigens, MAP3K7 in multiple sclerosis and NRXN1 in narcolepsy.

Paper IV: Epitopes of anti-RIFIN antibodies and characterization of *rif*-expressing *Plasmodium falciparum* parasites by RNA sequencing

The protozoan parasite *Plasmodium falciparum* is responsible for causing the most prevalent and severe form of malaria in the world [349]. The protein family, repetitive interspersed families of polypeptides (RIFINs) constitute the largest family of variable surface antigens on the surfaces of infected red blood cells [350]. In addition, it has been shown that the RIFINs have an important role in mediating blood group A rosetting of infected red blood cells, which leads to sequestration of the parasite. Since rosetting is a very important denominator for the severity of *Plasmodium falciparum* malaria [351], it is of great interest to study the expression of RIFINs. In this study, three different polyclonal anti-RIFIN antibodies were epitope mapped using high-density peptide microarrays in order to relate their epitopes to functional performance in different assays when analyzing three rosetting strains and five non-rosetting strains of *Plasmodium falciparum*.

Initially, 11 different anti-RIFIN IgG antibodies had been generated in 10 rabbits and 1 goat with different immunogens representing different RIFIN proteins. When performing western blot of the eight different lysates, three out of the 11 anti-RIFIN antibodies, revealed a band corresponding to the predicted molecular weight of RIFINs of approximately 35 kDa. All three functional antibodies in western blot were raised against the RIFIN, PF3D7_0100400. The first antibody was generated towards a C-terminal peptide in rabbit (Rα-RIF_C), the second towards a conserved indel peptide Rα-RIF_I in rabbit, and the third towards the full-length PF3D7_0100400 (Gα-RIF) in goat. Most of these gave single bands in the analyzed lysates, except for the Gα-RIF that in addition to giving bands of predicted molecular weight, also reacted with unknown components of high molecular weight.

When performing immunofluorescence microscopy of infected red blood cells only one of the antibodies, the antibody raised against the C-terminal region of PF3D7_0100400 (Rα-RIF_C), was able to bind the S1.2R parasite (well-studied rosetting parasite with RIFIN expression). Furthermore, this antibody could stain four other strains, where half of them were not rosetting strains. By also

including an antibody towards the Maurer's cleft, it could be seen that as in contrary to previous data, the RIFINs do not seem to be exported through the Maurer's cleft.

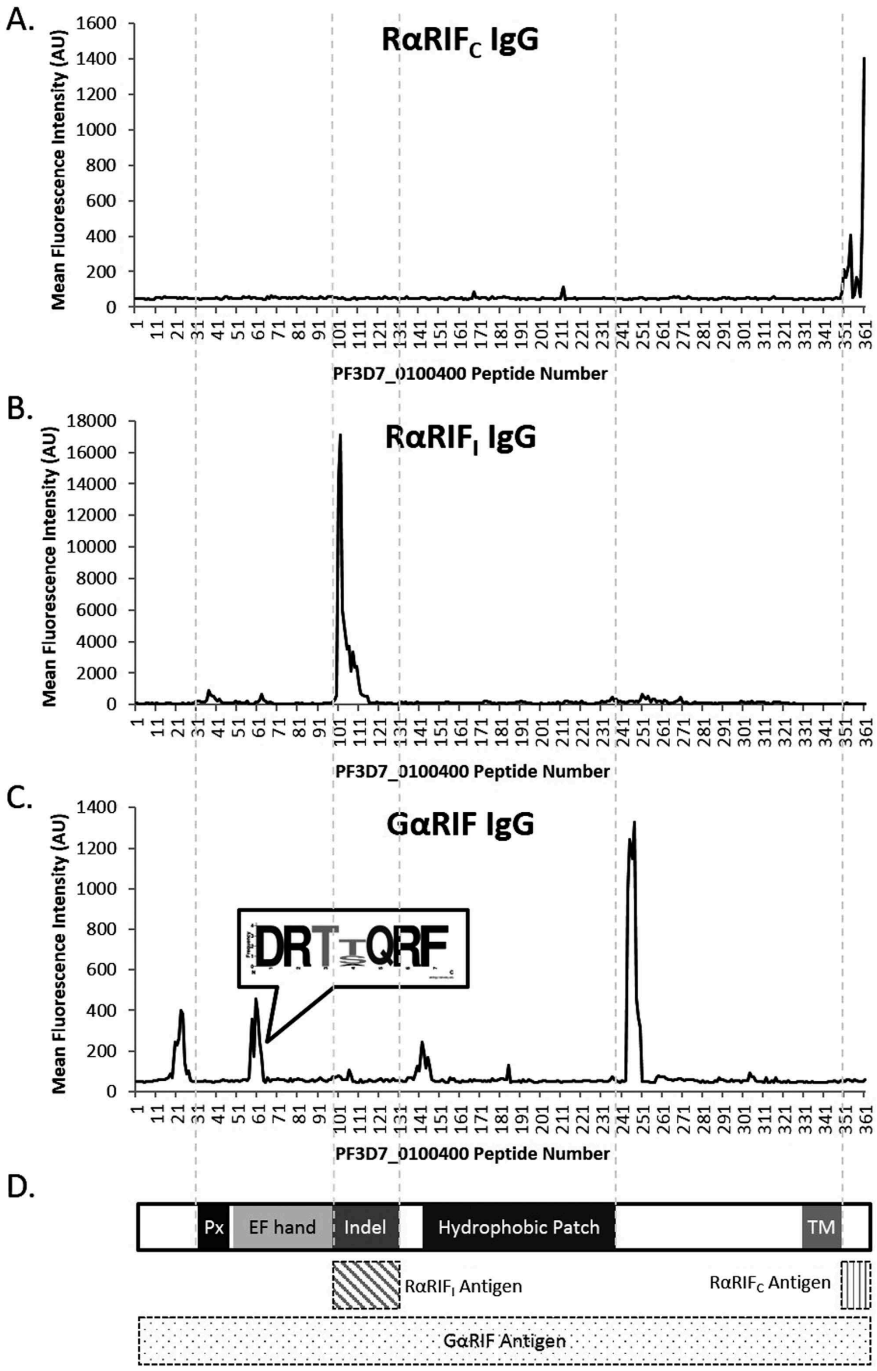


***Figure 13**. Epitope mapping results from the peptide microarrays. A) Epitopes for Rα-RIF$_C$ that were immunized with a C-terminal antigen. B) Epitopes for Rα-RIF$_I$, the antibody immunized with the indel antigen, C) Epitopes for Gα-RIF that were immunized with the full-length protein. D) Schematic overview of the PF3D7_0100400 protein.*

The antibodies were epitope mapped on high-density peptide microarrays consisting of 175,000 12-mer peptides with 11 amino acid overlap, covering most of the known surface proteins of *Plasmodium falciparum.* The antibody Rα-RIF$_C$ showed reactivities towards peptides representing half of the included RIFINs, exclusively interacting with the C-terminal (figure 13), thus showing it is in line with the performance on western blot as well as immunofluorescence microscopy. The antibody raised against the indel-region, Rα-RIF$_I$, showed reactivity towards a third of all included RIFINs on the peptide microarray, binding to peptides not surprisingly located in the indel-region (figure 13). This was in contrast to the immunofluorescence microscopy where it did not give any signal when analyzed with the different strains. Furthermore, the indel-region of the antigen is predicted to be expressed extracellularly of the infected red blood cells and therefore an explanation could be that the continuous epitopes recognized on the peptide arrays, might not be accessible in presence of the full-length protein, and other proteins. For the antibody raised against the full-length RIFIN PF3D7_0100400, cross-reactivities could be seen to other surface antigens than the RIFINs that were represented on the peptide microarray. This might also explain that why several bands of higher molecular weights could be observed in the western blot.

In order to validate these findings further, four strains *Plasmodium falciparum* strains were selected and analyzed by RNA sequencing, to investigate the transcript levels of RIFINs. For two of the strains FCR3CSA and IT4CD36ICAM1, where the antibody Rα-RIF$_C$ gave signal in both western blot and immunofluorescence microscopy, two highly expressed *rif*-transcripts were identified, PFIT_bin05750 and PFIT_0835500 respectively. On the other hand, the NF54CSA and S1.2NR strains, which the antibody Rα-RIF$_C$ did not react with in western blot or immunofluorescence, had unlike the other two strains, no highly expressed *rif*-transcript.

In summary, purified polyclonal antibodies raised against the protein family RIFINs that are found on the surface of infected red blood cells, were epitope mapped on a peptide microarray of 175,000 unique peptides covering most surface proteins of *Plasmodium falciparum.* The epitopes of the antibodies were compared to performance in western blot and immunofluorescence microscopy. Furthermore, RIFINs have been suggested to be involved in the rosetting mechanisms, in which infected red blood cells can escape the host's immune

response. By combining the results of the characterized epitopes of the antibodies, together with performance in western blot and immunofluorescence microscopy, and an orthogonal validation of transcript expressions through RNA sequencing; two previously uncharacterized non-rosetting parasite strains were found to express RIFINs.

Challenges and considerations

The field of proteomics has evolved tremendously in the last decades, which have enabled analysis of thousands of proteins in parallel. This has constituted an important part of discovering proteins that could be of clinical relevance, i.e. biomarkers. In the following sections, the analytical challenges in identifying the autoantigens presented in this thesis are discussed.

The starting points in **Paper I - Paper IV** have been to either use high-density peptide arrays or protein fragment arrays to profile antibodies. Although these technologies provide a platform for large-scale analysis of antibodies and potential autoantibodies, a valid concern is that a portion of disease-relevant reactivities might go undiscovered. The protein fragments (PrESTs) that are spotted on the arrays are in general between 50-150 amino acids, with in average around 80 amino acids, representing only unique sequences of each protein-coding gene. As a consequence, potential autoantibodies targeting the non-unique portions of a protein will go undetected, as no protein fragment is generated towards such regions in the Human Protein Atlas project (with a few exceptions). However, the disease relevance of such antibody can be questioned as they can potentially demonstrate cross-reactivity between non-unique regions of at least two protein-coding genes. Furthermore, despite the protein fragments are around 80 amino acids, the native conformation will likely not be assumed, and therefore antibodies that are directed towards discontinuous epitopes might go undetected.

In **Paper III - Paper IV**, high-density peptide microarrays were used, where the array consisted of overlapping 12-mer peptides representing a large set of proteins. As discontinuous epitopes in average span between 50-79 amino acids in the primary structure [220], antibodies directed to discontinuous epitopes would have gone undetected. However, as discussed in Chapter 4, antibodies that bind to discontinuous epitopes can be detected on shorter peptides

through mimotopes [298]. Also, discontinuous epitopes consist of several stretches of continuous residues with an average maximum length of 5 amino acids [212]. If provided sufficient binding strength of the antibodies, such antibodies can be detected. On the high-density peptide microarrays, in contrast to protein fragment arrays, the relation between peptide and corresponding protein is not as clear. Due to sequence similarities between the protein-coding genes, individual peptides can be mapped to multiple proteins, which makes it difficult to e.g. identify the actual autoantigen. In addition, sequence similarities between proteins on peptide level are also present between organisms, and therefore peptides derived from pathogens can be represented by human peptides. By comparing peptide sequences using basic local alignment search tool (BLAST) [352], sequence similarities to e.g. bacterial and viral proteins can be found. As a consequence of these above-mentioned ambiguities, despite of observed antibody reactivities towards antigens or peptides representing self, it cannot be ruled out that the reactivities are due to cross-reactivity. Since molecular mimicry [255] is one of the hypothesized mechanism of autoimmunity, such cross-reactivities may still be valuable to study.

Another challenge of the presented work is the heterogeneity in antibody reactivity observed between individuals. As seen in **Paper I - Paper III**, most of the reactivities towards the peptides and proteins were only observed in single individuals, which are in line with previous work [110,353]. Given the immense capability of the immune system to create antibodies with a wide range of specificities, these observations are not unlikely. In addition, it cannot be ruled out that the heterogeneity observed between individuals diagnosed with the same disease could potentially be due to detection of cross-reactive anti-pathogenic antibodies. Also, given the multifactorial nature of autoimmune diseases and the various hypothesized disease etiologies [242,255], it is not unexpected to observe such heterogeneity. Not only is it a large inter-individual heterogeneity; but also an intra-individual heterogeneity that can be observed over time as previously shown [354,355]. This enormous heterogeneity in the antibody repertoire require adequate number of samples to be analyzed in order to reveal patterns that could indicate certain autoreactivity in a disease. However, there is a tradeoff between the sample throughput and the multiplexing capacity that needs to be considered. By initially being as inclusive as possible in the study selection of peptides or protein fragment targets, the number of analyzed will consequently be limited in following verifications on larger sample collections as seen in **Paper I** and **Paper III**.

The initial screenings are performed on small sample groups to enable as many analytes as possible to be analyzed, and given the heterogeneous character of the data, there have been challenges in filter which reactivities are relevant to further study and which are not.

Concluding remarks

Throughout the chapters in this thesis I have provided a glimpse of the immense capabilities of the immune system in general and antibodies in particular, and how this class of molecules can generate seemingly endless specificities in order to protect the host from pathogenic substances. This ability has directly or indirectly been utilized for design of vaccines, diagnostics of infectious diseases, therapeutic antibodies, but also used to study proteins, one of the major building blocks of life. However, in some cases the immune system can fail in distinguishing between self and non-self and thus give rise to an autoimmune disease and circulating autoantibodies. Given the polyspecific nature of antibodies, their role in disease, but also as being an important component in the molecular toolbox, there are a lot of efforts in trying to understand what antibodies bind to and can bind to. The work presented in this thesis aims at profiling antibodies and to characterize their antigens, as well as epitope mapping to gain a deeper insight in the specificities. In **Paper I** - **Paper III**, the antibody repertoire in plasma and serum samples from patients suffering from autoimmune diseases were profiled in order to characterize potential autoantigens and autoantibodies. In **Paper IV** purified polyclonal antibodies were characterized to relate their specificities to function.

This thesis covers historical aspects of proteomics, the field of large-scale analysis of proteins. It introduces the different methods to study and detect antibodies, as antibodies play an important role in health and disease in organisms and constitute a tool for studying proteins. Since the immune system can in some diseases turn itself against the host and thus cause an autoimmune reaction including production of autoantibodies, a common way to identify and characterize what the immune system is targeting, is to profile the antibody repertoire.

Despite the efforts in understanding the pathological mechanisms of the suggested autoimmune disease narcolepsy, the etiology remains unknown [356]. Therefore the potential autoantigens that were identified in **Paper I** might be

of interest for increased understanding and further characterization of narcolepsy. One of the identified autoantigens, NT5C1A, has been previously described in the suggested autoimmune disease sporadic inclusion body myositis [341]. Because of the different phenotypes between those diseases, it is interesting to further study what the autoreactivities may be caused by, and if it may be due to cross-reactive antibodies or a shared autoreactivity that is clinically relevant. In addition, in **Paper III**, a peptide representing NRXN1 was identified to have higher frequency of antibody reactivities in patients with narcolepsy. NRXN1 is a well-studied neuronal protein, where mutations on a gene level are associated to different neuropsychiatric and developmental disorders, including schizophrenia and autism [357]. Although its association to narcolepsy has not previously been studied, NRXN1 has been suggested to be involved in sleep dysregulation [358]. Whether aberrations of NRXN1, due to autoantibodies or mutations, have a role in narcolepsy remains to be seen and will require further work.

In **Paper II**, antibody repertoires in patients with first-episode psychosis were profiled on human protein fragments to find potential autoantibodies. A key motivation for this study was that there is a lack of biomarkers that can be used for patients with psychiatric disorders including schizophrenia, since such analyses are almost solely based on behavioral characteristics and self-reporting [359]. Therefore, the identification of the potential autoantibodies towards the N-terminal portion of the three proteins PAGE2B, PAGE2 and PAGE5 could potentially be valuable for clinical use, as patients can be identified preferentially at early onset and get suitable treatment. Although, there is a risk that the autoantibodies are cross-reactive since the PAGE proteins are tumor-associated and mainly expressed in testis and ovaries [360], the potential usefulness may still be unaffected.

Furthermore, in **Paper III** potential autoantibodies towards peptides representing MAP3K7 was observed in patients with multiple sclerosis. Although multiple sclerosis is regarded as a T-cell-driven autoimmune disease, the B-cell involvement has not been as well-studied [361]. The potential autoantigen MAP3K7 is a protein involved in activation of inflammation pathways and suggested to contribute to the demyelination processes in the animal model of multiple sclerosis, EAE. Therefore, the finding of anti-MAP3K7 antibodies in a subgroup of multiple sclerosis patients could be valuable for the further understanding of the disease. Additionally, it can

potentially facilitate treatment decisions of patients, as it is believed that there are subgroups in multiple sclerosis that are not yet characterized [361].

As seen in **Paper I - Paper III**, microarray technologies have enabled identification of potential autoantigens and epitopes associated to disease. The next steps to elucidate the role of the potential autoantigens presented in **Paper I - Paper III** include analyzing more patient samples, and preferentially also cerebrospinal fluid due to its proximity to the brain, as psychosis, narcolepsy and multiple sclerosis affect the brain. In addition, immunohistochemical analyses on brain tissue from purified autoantibodies could further increase the understanding of the target of these autoantibodies in the respective patients.

A major hurdle of the microarray analysis has been to deal with the individual heterogeneity seen in antibody reactivities [110,353]. Due to the heterogeneity of antibodies between individuals, a potential solution to this could be to monitor and study autoantibody repertoires over time to be able to better examine and monitor changes that can be linked to medical phenotypes [354]. Such large-scale analysis of the antibody repertoires has recently been reported by Stafford *et al.* in 2016 [355], and shown to be a valuable tool in linking the observed immunosignatures to immunological changes, and thus potentially disease.

As seen in **Paper IV**, antibodies that are generated towards the same protein can perform differently in the assays they are used in. Also through the advances of microarray technologies, large-scale epitope mapping of the antibodies on peptide microarrays, consequently enabled relating their epitopes to performance in the different assays. Similar observations have also been reported in a study by Hjelm *et al.* in 2011, where epitope-specific antibodies showed different functionality in different immunoassays [202]. In addition, the antibody performances were validated using an antibody-independent method, mRNA sequencing. The presented strategies are in line with the recent proposal for antibody validation [133]. As validated antibodies constitute the foundation in reproducible research in molecular biology, it is important that the research groups that use antibodies perform adequate validation of antibodies to ensure that they recognize the intended target.

Although three-dimensional structures of antibodies and antigens suggest that epitopes are exclusively discontinuous [212,220], the tools for mapping discontinuous epitopes have been limited. Instead focus has historically been on studying short continuous stretches of amino acid segments. These continuous epitopes represent a part from the discontinuous epitope, which provides enough binding strength to be detected. As we have seen, there are many different methods and approaches for studying such continuous epitopes. A very promising approach is the use of random peptide microarray designs, in which most potential tetramers (83%) and 27% of all potential pentamers can be synthesized on the microarray surface [301]. Through this method, the combined advantages of surface display-based methods, i.e. the generation of large libraries of probes, along with the simplicity and speed of analyzing microarray features, enables analysis of almost all potential continuous epitopes. Furthermore, the random peptide microarray can also enable mapping of discontinuous epitopes through mimotopes.

Another promising approach that has been discussed in this thesis is the surface display-based methods. These have in the last decade undergone tremendous improvements, much thanks to the development of next-generation sequencing, but also library preparations enabling hundreds of thousands unique clones to be generated [319]. An apparent advantage over protein microarrays is that there is no need for purification, and once the sequence library is generated it can easily be scaled up, and therefore they are also cheaper than protein microarrays [319]. As cost of sequencing steadily decreases [362], surface display-based methods for studying antigens and epitopes will become even more favorable.

Yet another very interesting potential of analyzing antibodies and their antigens and epitopes are methods based on mass-spectrometry. One approach introduced by Merl *et al.* in 2013 [363], used mass-spectrometry to analyze autoantigens in autoimmune uveitis, by enriching immune complexes, from 10 mL vitrectomy samples from horses, on a solid surface. Although the large sample volume seem highly impractical for analysis of human body fluids, it can be assumed that blood-based body fluids contain higher levels of antibodies. Furthermore, mass spectrometry-based methods have advanced enormously in the last years, and especially in terms of sensitivity, enabling analysis of low abundant molecules, and therefore it might be feasible to use a similar approach on human samples. By combining the setup of Merl *et al.* [363], with the addition of protein mixtures from cell and tissue lysates, it will

be possible to analyze specific sets of antigens, analogous to the reverse capture autoantibody microarray described by Qin *et al.* in 2006 [364]. Given adequate sensitivity of the system, it might also be possible to combine such analysis with epitope mapping strategies as previously discussed [284] and thus making mass spectrometry-based methods very powerful tools for studying antibodies and their interaction partners.

Two of the driving forces within proteomics have been to be able to analyze more samples, higher throughput, and to analyze as many proteins as possible in parallel, hence higher multiplexing and increasing the degree of analytical depth [365-367]. Given the technological advances, there are now an abundance of methods that fulfill these criteria for characterizing and studying antibodies, their antigens and epitopes in proteome scale [320,368], including increasingly reliable tools to predict epitopes. In addition to analyzing antibodies to understanding disease etiology of immune diseases as well as surveying the immune response to foreign and self-epitopes, there have been several recent publications on mapping the peptidome of MHC molecules using mass spectrometry-based methods [369-372]. These analyses will in conjunction to analysis of antibodies and their epitopes, complement each other in order to get a comprehensive picture of the immune system in health and disease. Altogether, the field of epitope mapping and antibody profiling can facilitate the development of drugs as well as vaccines, and ultimately contribute to an increased understanding of autoimmune diseases as well as infectious diseases.

Acknowledgements

I would like to acknowledge the financial support from ProNova VINN Excellence Centre for Protein Technology, AstraZeneca-Karolinska Institutet Joint Research Program in Translational Science and KTH Center of Applied Precision Medicine funded by the Erling-Persson Familiy Foundation, which have enabled the work presented in this thesis.

I would like to thank all co-authors and colleagues that have helped me to accomplish the work presented in this thesis, but also ongoing work and published work that are not included in this thesis.

My supervisors, **Peter**, **Maja** and **Jochen**, thank you so much! I would not be able to accomplish this without you.

Peter, tack för att jag fick göra mitt exjobb hos dig och sedan fortsätta som doktorand. Tack för all engagemang, uppmuntran och vägledning i projekten. Du har delat med dig av otroligt mycket klokheter under dessa fem år som jag alltid kommer ha med mig! **Maja**, tack för att du fick mig att få upp ögonen för proteomik i samband med kex-jobbet för drygt 7 år sedan. Tack för all handledning, stöd och hjälp under dessa år! **Jochen**, thank you for the help and inspiration on the different projects, in particular the protein profiling projects.

I would like to thank everyone involved in the **Human Protein Atlas** project and **Atlas Antibodies**, for creating the resource of antibodies and protein fragments used in the presented papers, but also enabled proteome-wide human atlases. **Mathias**, särskilt stort tack till dig som har möjliggjort HPA-projektets framgångar.

I would like to thank **all current** and **former** members of the **PAPP group**, **Alfa2**, and associates for all help and support in the different projects, and for creating a pleasant and productive working atmosphere. In particular I would like to thank:

Björn, tack för att du introducerade mig till peptidarrayerna och varit så hjälpsam i de experimenten. Tack också för allt härligt häng, alla fantastiska sushi-luncher och lånet av bryggverket! **Elin**, tack för alla gånger jag fick hjälp av dig på labbet i de olika projekten och för att du är en sån härlig glädjespridare. Tack också för alla skojiga resor med jobbet, men också utanför! **Fredrik**, i skrivande stund så var det exakt 10 år sedan vi började på KTH. Det är så kul att vi båda hamnade inom forskningen och båda till slut även på SciLifeLab efter alla dessa år. Tack för allt skoj på jobbet och utanför! Glöm inte, om forskningen skiter sig så kan vi alltid starta ett mikrobryggeri tillsammans med **David R**! David R, tack **Sanna**, tack för alla lärorika diskussioner kring data-analys, antikroppar, och olika immunoassays. **Frida**, tack för att du delat med dig av dina idéer och funderingar och fått mig att utveckla mina R-kunskaper. Tack även för det härliga resesällskapet i bl.a. Brixen och Vancouver. **Hammou**, tack för alla intressanta diskussioner och för allt härligt häng på och utanför jobbet! **Anna**, tack för all hjälp med narkolepsiprojekten. **Ulrika**, tack för all hjälp med de olika peptidarray-projekten. **Devin**, thank you for the nice board game nights, but also for intresting discussions about machine learning. **Christian,** tack för allt härligt sällskap på jobbet och utanför. **David J**, thank you for being such a nice office neighbor and also for understanding my frustration in writing a certain manuscript. **Maria M**, tack för all hjälp på labbet och med de olika projekten och för att du alltid lyssnar och låter mig störa dig för att bolla idéer eller bara ventilera. **Mun-Gwan**, thank you for helping me out with all type of statistical questions in the different projects. Also, thank you for inspiring me to learn more on programming. **Ronald**, tack för all hjälp med PrEST-arrayerna. **Kimi**, tack för all hjälp med plasma profiling och alla härliga lunchdiskussioner med bl.a. **Hammou**. **Eni**, stort tack för all hjälp med PrEST:arna i psykos- och narkolepsiprojekten. **Sofia**, tack för att jag fick handleda dig under ditt exjobb, och att du tog dig an det kluriga fibbe-projektet! **Evelina**, tack för all hjälp med psykos-projektet. **Burcu**, thank you for all effort and help in the different projects. **Cecilia H** och **Cecilia M**, stort tack för all hjälp med SLE-projekten och tack **Cecilia H** för hjälp i psykos-

och narkolepsi-projekten. **Frauke**, thank you for help with the different projects and teaching me vital conference presentation skills. **Maria-Jesus**, thank you for helping in the lab with the sandwich assays. **Claudia**, thank you for helping me out in the SLE and fibulin1 projects, and for answering all my questions on mass spectrometry. **Laura**, thank you for all the help in the lab. **Elisa,** thank you for the help on autoimmunity projects. **Lucía**, thank you for working on figuring out how to deal with the rheumatoid factors! I also would like to thank the rest of the **PAPP** group, including **Annika**, **Gabriella**, **Jennie**, **Julia**, **Matilda**, **Ragna,** and **Tea**.

Tack även till hela Taipei-gänget, och särskilt till **Sanna**, **Marie U**, **Evelina**, **Fredrik** och **Ulrika** för en oförglömlig resa, bl.a. innehållandes "no laws" samt en utflykt till nordöstra udden.

Tack även till **HPA-IT**: **Fredric**, **Martin**, **Kalle**, **Lukas**, **Per** och andra. Utan er hade vår forskning varit betydligt svårare att genomföra.

Thank you **David R**, **Erik**, **Megan**, **Sanna**, **Frida**, **Björn**, **Fredrik**, **and Maria M** for proofreading parts of the thesis.

Jag vill även tacka min familj, **Mamma**, **Pappa** och min tvillingbror, **Aidin**. Tack för att ni alltid har stöttat mig i mina studier, alltid ställt upp och uppmuntrat mig i alla situationer!

Sara, tack för all motivation, stöd, förståelse och tålamod under det senaste året! Du är bäst, jag älskar dig! <3

References

1. McKay, C.P. (2004) What is life--and how do we search for it in other worlds? *PLoS Biol*, **2**, E302.
2. Koshland, D.E., Jr. (2002) Special essay. The seven pillars of life. *Science*, **295**, 2215-2216.
3. Hooke, R. and Lessing J. Rosenwald Collection (Library of Congress). (1665) *Micrographia: or, Some physiological descriptions of minute bodies made by magnifying glasses. With observations and inquiries thereupon*. Printed by J. Martyn and J. Allestry, London,.
4. Lane, N. (2015) The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'. *Philos T R Soc B*, **370**.
5. Schwann, T., Smith, H. and Schleiden, M.J. (1847) *Microscopical researches into the accordance in the structure and growth of animals and plants*. The Sydenham Society, London.
6. Fourcroy, A.-F.o.d. (1794) *Élémens d'histoire naturelle et de chimie*. 5. éd.; ed. Cuchet, Paris.
7. Tanford, C. and Reynolds, J.A. (2001) *Nature's robots : a history of proteins*. Oxford University Press, Oxford ; New York.
8. Vickery, H.B. (1950) The origin of the word protein. *Yale J Biol Med*, **22**, 387-393.
9. Miquel, E.A.W., Mulder, G.J. and Wenckebach, W. (1838) *Bulletin des sciences physiques et naturelles en Néerlande*, Leyde, Rotterdam.
10. Darwin, C. (1859) *On the origin of species by means of natural selection*. J. Murray, London,.
11. Winther, R.G. (2000) Darwin on variation and heredity. *J Hist Biol*, **33**, 425-455.
12. Bateson, W. (2010) *Mendel's principles of heredity*. Dover ed. Dover Publications, Mineola, N.Y.
13. Johannsen, W. (2014) The genotype conception of heredity. 1911. *Int J Epidemiol*, **43**, 989-1000.
14. Wanscher, J.H. (1975) An analysis of Wilhelm Johannsen's genetical term "genotype' 1909--26. *Hereditas*, **79**, 1-4.
15. Johannsen, W. (1909) *Elemente der exakten erblichkeitslehre*. . G. Fischer, Jena, Germany.
16. Dawkins, R. (2016) *The extended phenotype : the long reach of the gene*. Oxford University Press, Oxford, United Kingdom.
17. Dawkins, R. (1976) *The selfish gene*. Oxford University Press, New York.
18. Danchin, E., Charmantier, A., Champagne, F.A., Mesoudi, A., Pujol, B. and Blanchet, S. (2011) Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet*, **12**, 475-486.
19. Dahm, R. (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet*, **122**, 565-581.
20. Olby, R.C. (1994) *The path to the double helix : the discovery of DNA*. Dover Publications, New York.
21. Sutton, W.S. (1903) The Chromosomes in Heredity. *Biological Bulletin*, **4**, 231-251.
22. Sutton, W.S. (1902) On the morphology of the chromosome group in Brachystola magna. *Biological Bulletin*, **4**, 24-39.
23. Van Holde, K.E. (1989) *Chromatin*. Springer-Verlag, New York.
24. Levene, P.A. and London, E.S. (1929) The structure of thymonucleic acid. *J. Biol. Chem.*, **83**, 793-802.
25. Levene, P.A. and Mori, T. (1929) Ribodesose and xylodesose and their bearing on the structure of thyminose. *J. Biol. Chem.*, **83**, 803-816.
26. Levene, P.A. (1938) Structure of desoxyribonucleic acid. *J. Biol. Chem.*, **126**, 63-66.
27. Griffith, F. (1928) The Significance of Pneumococcal Types. *J Hyg (Lond)*, **27**, 113-159.

28. Avery, O.T., Macleod, C.M. and McCarty, M. (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine*, **79**, 137-158.
29. Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, **6**, 201-209.
30. Hershey, A.D. and Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol*, **36**, 39-56.
31. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
32. Crick, F.H. (1958) On protein synthesis. *Symp Soc Exp Biol*, **12**, 138-163.
33. Astrachan, L. and Volkin, E. (1958) Properties of ribonucleic acid turnover in T2-infected Escherichia coli. *Biochim Biophys Acta*, **29**, 536-544.
34. Brenner, S., Jacob, F. and Meselson, M. (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, **190**, 576-581.
35. Simoni, R.D., Hill, R.L. and Vaughan, M. (2002) The discovery of the amino acid threonine: the work of William C. Rose [classical article]. *J Biol Chem*, **277**, E25.
36. Tiselius, A. (1937) A new apparatus for electrophoretic analysis of colloidal mixtures. *Transactions of the Faraday Society*, **33**, 524-531.
37. Pauling, L., Itano, H.A. and et al. (1949) Sickle cell anemia a molecular disease. *Science*, **110**, 543-548.
38. Sanger, F. (1988) Sequences, sequences, and sequences. *Annu Rev Biochem*, **57**, 1-28.
39. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M. and Smith, M. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, **265**, 687-695.
40. Costantini, F. and Lacy, E. (1981) Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature*, **294**, 92-94.
41. Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**, 1350-1354.
42. Watson, J.D. and Jordan, E. (1989) The Human Genome Program at the National Institutes of Health. *Genomics*, **5**, 654-656.
43. Collins, F. and Galas, D. (1993) A new five-year plan for the U.S. Human Genome Project. *Science*, **262**, 43-46.
44. Yadav, S.P. (2007) The wholeness in suffix -omics, -omes, and the word om. *J Biomol Tech*, **18**, 277.
45. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith,

D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J. and International Human Genome Sequencing, C. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

46. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. and Zhu, X. (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.

47. Gilbert, W. (1991) Towards a paradigm shift in biology. *Nature*, **349**, 99.

48. Kell, D.B. and Oliver, S.G. (2004) Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays*, **26**, 99-105.
49. Gerlai, R. (1996) Gene-targeting studies of mammalian behavior: is it the mutation or the background genotype? *Trends Neurosci*, **19**, 177-181.
50. Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.
51. Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484-487.
52. Patterson, S.D. and Aebersold, R.H. (2003) Proteomics: the first decade and beyond. *Nature genetics*, **33 Suppl**, 311-323.
53. Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, **19**, 1853-1861.
54. Tyers, M. and Mann, M. (2003) From genomics to proteomics. *Nature*, **422**, 193-197.
55. Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I. and Overington, J.P. (2017) A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*, **16**, 19-34.
56. Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198-207.
57. Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Bjork, L., Breckels, L.M., Backstrom, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey, C.M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J.M., Sivertsson, A., Sjostedt, E., Skogs, M., Stadler, C., Sullivan, D.P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen, M., Mardinoglu, A., Ponten, F., von Feilitzen, K., Lilley, K.S., Uhlen, M. and Lundberg, E. (2017) A subcellular map of the human proteome. *Science*, **356**.
58. Reif, D.M., White, B.C. and Moore, J.H. (2004) Integrated analysis of genetic, genomic and proteomic data. *Expert Rev Proteomics*, **1**, 67-75.
59. Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. and Anker, P. (2001) About the possible origin and mechanism of circulating DNA - Apoptosis and active DNA release. *Clin Chim Acta*, **313**, 139-142.
60. Lewis, J.M., Heineck, D.P. and Heller, M.J. (2015) Detecting cancer biomarkers in blood: challenges for new molecular diagnostic and point-of-care tests using cell-free nucleic acids. *Expert review of molecular diagnostics*, **15**, 1187-1200.
61. van der Vaart, M. and Pretorius, P.J. (2007) The origin of circulating free DNA. *Clin Chem*, **53**, 2215.
62. Garcia, V., Garcia, J.M., Pena, C., Silva, J., Dominguez, G., Lorenzo, Y., Diaz, R., Espinosa, P., de Sola, J.G., Cantos, B. and Bonilla, F. (2008) Free circulating mRNA in plasma from breast cancer patients and clinical outcome. *Cancer Lett*, **263**, 312-320.
63. Narita, M., Saito, A., Kojima, A., Iwabuchi, M., Satoh, N., Uchiyama, T., Yamahira, A., Furukawa, T., Sone, H. and Takahashi, M. (2012) Quantification of BCR-ABL mRNA in plasma/serum of patients with chronic myelogenous leukemia. *Int J Med Sci*, **9**, 901-908.
64. Thierry, A.R., El Messaoudi, S., Gahan, P.B., Anker, P. and Stroun, M. (2016) Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev*, **35**, 347-376.
65. Kishikawa, T., Otsuka, M., Ohno, M., Yoshikawa, T., Takata, A. and Koike, K. (2015) Circulating RNAs as new biomarkers for detecting pancreatic cancer. *World J Gastroentero*, **21**, 8527-8540.
66. Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*, **13**, 227-232.
67. de Sousa Abreu, R., Penalva, L.O., Marcotte, E.M. and Vogel, C. (2009) Global signatures of protein and mRNA expression levels. *Mol Biosyst*, **5**, 1512-1526.
68. Ning, K., Fermin, D. and Nesvizhskii, A.I. (2012) Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *Journal of proteome research*, **11**, 2261-2271.

69. Gry, M., Rimini, R., Stromberg, S., Asplund, A., Ponten, F., Uhlen, M. and Nilsson, P. (2009) Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics*, **10**, 365.

70. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337-342.

71. Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., Algenas, C., Lundeberg, J., Mann, M. and Uhlen, M. (2010) Defining the transcriptome and proteome in three functionally different human cell lines. *Mol Syst Biol*, **6**, 450.

72. Edfors, F., Danielsson, F., Hallstrom, B.M., Kall, L., Lundberg, E., Ponten, F., Forsstrom, B. and Uhlen, M. (2016) Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol*, **12**, 883.

73. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.H., Bantscheff, M., Gerstmair, A., Faerber, F. and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature*, **509**, 582-587.

74. Brändén, C.-I. and Tooze, J. (1999) *Introduction to protein structure*. 2nd ed. Garland Pub., New York.

75. Petsko, G.A. and Ringe, D. (2004) *Protein structure and function*. New Science Press, London.

76. Berg, J.M., Tymoczko, J.L., Stryer, L. and Stryer, L. (2007) *Biochemistry*. 6th ed. W.H. Freeman, New York.

77. Pauling, L., Corey, R.B. and Branson, H.R. (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, **37**, 205-211.

78. Lockless, S.W. and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, **286**, 295-299.

79. Levitt, M. and Chothia, C. (1976) Structural Patterns in Globular Proteins. *Nature*, **261**, 552-558.

80. Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E. and Ben-Tal, N. (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics*, **19**, 163-164.

81. Demarest, S.J. and Glaser, S.M. (2008) Antibody therapeutics, antibody engineering, and the merits of protein stability. *Curr Opin Drug Disc*, **11**, 675-687.

82. Vonderosten, C., Branner, S., Hastrup, S., Hedegaard, L., Rasmussen, M.D., Bisgardfrantzen, H., Carlsen, S. and Mikkelsen, J.M. (1993) Protein Engineering of Subtilisins to Improve Stability in Detergent Formulations. *J Biotechnol*, **28**, 55-68.

83. Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Giron, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R. and Flicek, P. (2016) Ensembl 2016. *Nucleic Acids Res*, **44**, D710-716.

84. Ponomarenko, E.A., Poverennaya, E.V., Ilgisonis, E.V., Pyatnitskiy, M.A., Kopylov, A.T., Zgoda, V.G., Lisitsa, A.V. and Archakov, A.I. (2016) The Size of the Human Proteome: The Width and Depth. *Int J Anal Chem*, **2016**, 7436849.

85. Engvall, E. and Perlmann, P. (1971) Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry*, **8**, 871-874.

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

87. Stoevesandt, O. and Taussig, M.J. (2012) Affinity proteomics: the role of specific binding reagents in human proteome analysis. *Expert Rev Proteomic*, **9**, 401-414.

88. Uhlen, M., Graslund, S. and Sundstrom, M. (2008) A pilot project to generate affinity reagents to human proteins. *Nature Methods*, **5**, 854-855.

89. Taussig, M.J., Stoevesandt, O., Borrebaeck, C.A.K., Bradbury, A.R., Cahill, D., Cambillau, C., de Daruvar, A., Dubel, S., Eichler, J., Frank, R., Gibson, T.J., Gloriam, D., Gold, L., Herberg, F.W., Hermjakob, H., Hoheisel, J.D., Joos, T.O., Kallioniemi, O., Koegll, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muyldermans, S., Nygren, P.A., Palcy, S., Pluckthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D.J., Skerra, A., Templin, M., Ueffing, M. and Uhlen, M. (2007) ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature Methods*, **4**, 13-17.

90. Ge, H. (2000) UPA, a universal protein array system for quantitative detection of protein-protein, protein-DNA, protein-RNA and protein-ligand interactions. *Nucleic Acids Res*, **28**, e3.

91. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M. and Snyder, M. (2001) Global analysis of protein activities using proteome chips. *Science*, **293**, 2101-2105.

92. Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A. and Snyder, M. (2000) Analysis of yeast protein kinases using protein chips. *Nature genetics*, **26**, 283-289.

93. Fukui, S., Feizi, T., Galustian, C., Lawson, A.M. and Chai, W.G. (2002) Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nature biotechnology*, **20**, 1011-1017.

94. Wang, D.N., Liu, S.Y., Trummer, B.J., Deng, C. and Wang, A.L. (2002) Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature biotechnology*, **20**, 275-281.

95. MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science*, **289**, 1760-1763.

96. Haab, B.B. (2003) Methods and applications of antibody microarrays in cancer research. *Proteomics*, **3**, 2116-2122.

97. Ekins, R., Chu, F. and Biggart, E. (1989) Development of microspot multi-analyte ratiometric immunoassay using dual fluorescent-labelled antibodies. *Analytica Chimica Acta*, **227**, 73-96.

98. Rimini, R., Schwenk, J.M., Sundberg, M., Sjoberg, R., Klevebring, D., Gry, M., Uhlen, M. and Nilsson, P. (2009) Validation of serum protein profiles by a dual antibody array approach. *J Proteomics*, **73**, 252-266.

99. Schroder, C., Jacob, A., Tonack, S., Radon, T.P., Sill, M., Zucknick, M., Ruffer, S., Costello, E., Neoptolemos, J.P., Crnogorac-Jurcevic, T., Bauer, A., Fellenberg, K. and Hoheisel, J.D. (2010) Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-related antibodies. *Molecular & cellular proteomics : MCP*, **9**, 1271-1280.

100. Schwenk, J.M., Gry, M., Rimini, R., Uhlen, M. and Nilsson, P. (2008) Antibody suspension bead arrays within serum proteomics. *Journal of proteome research*, **7**, 3168-3179.

101. Neiman, M., Hedberg, J.J., Donnes, P.R., Schuppe-Koistinen, I., Hanschke, S., Schindler, R., Uhlen, M., Schwenk, J.M. and Nilsson, P. (2011) Plasma profiling reveals human fibulin-1 as candidate marker for renal impairment. *Journal of proteome research*, **10**, 4925-4934.

102. Ingvarsson, J., Wingren, C., Carlsson, A., Ellmark, P., Wahren, B., Engstrom, G., Harmenberg, U., Krogh, M., Peterson, C. and Borrebaeck, C.A.K. (2008) Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics*, **8**, 2211-2219.

103. Collett, J.R., Cho, E.J. and Ellington, A.D. (2005) Production and processing of aptamer microarrays. *Methods*, **37**, 4-15.

104. Lundberg, M., Eriksson, A., Tran, B., Assarsson, E. and Fredriksson, S. (2011) Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res*, **39**.

105. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H.,

Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. and Ponten, F. (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.

106. Agaton, C., Galli, J., Hoiden Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., Wurtz, D., Hoog, C., Lundeberg, J., Stahl, S., Ponten, F. and Uhlen, M. (2003) Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Molecular & cellular proteomics : MCP*, **2**, 405-414.

107. Uhlen, M., Bjorling, E., Agaton, C., Szigyarto, C.A., Amini, B., Andersen, E., Andersson, A.C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergstrom, K., Brumer, H., Cerjan, D., Ekstrom, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Bjorklund, M.G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundeberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Skollermo, A., Steen, J., Stenvall, M., Sterky, F., Stromberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Walden, A., Wan, J., Wernerus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L.L., Hober, S. and Ponten, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Molecular & cellular proteomics : MCP*, **4**, 1920-1932.

108. Nilsson, P., Paavilainen, L., Larsson, K., Odling, J., Sundberg, M., Andersson, A.C., Kampf, C., Persson, A., Al-Khalili Szigyarto, C., Ottosson, J., Bjorling, E., Hober, S., Wernerus, H., Wester, K., Ponten, F. and Uhlen, M. (2005) Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics*, **5**, 4327-4337.

109. Berglund, L., Bjorling, E., Oksvold, P., Fagerberg, L., Asplund, A., Szigyarto, C.A., Persson, A., Ottosson, J., Wernerus, H., Nilsson, P., Lundberg, E., Sivertsson, A., Navani, S., Wester, K., Kampf, C., Hober, S., Ponten, F. and Uhlen, M. (2008) A genecentric Human Protein Atlas for expression profiles based on antibodies. *Molecular & cellular proteomics : MCP*, **7**, 2019-2027.

110. Ayoglu, B., Haggmark, A., Khademi, M., Olsson, T., Uhlen, M., Schwenk, J.M. and Nilsson, P. (2013) Autoantibody profiling in multiple sclerosis using arrays of human protein fragments. *Molecular & cellular proteomics : MCP*, **12**, 2657-2672.

111. Haggmark-Manberg, A., Zandian, A., Forsstrom, B., Khademi, M., Lima Bomfim, I., Hellstrom, C., Arnheim-Dahlstrom, L., Hallbook, T., Darin, N., Lundberg, I.E., Uhlen, M., Partinen, M., Schwenk, J.M., Olsson, T. and Nilsson, P. (2016) Autoantibody targets in vaccine-associated narcolepsy. *Autoimmunity*, **49**, 421-433.

112. Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature*, **405**, 837-846.

113. Nadler, W.M., Waidelich, D., Kerner, A., Hanke, S., Berg, R., Trumpp, A. and Rosli, C. (2017) MALDI versus ESI: The Impact of the Ion Source on Peptide Identification. *Journal of proteome research*, **16**, 1207-1215.

114. Mallick, P. and Kuster, B. (2010) Proteomics: a pragmatic perspective. *Nature biotechnology*, **28**, 695-709.

115. Frese, C.K., Altelaar, A.F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A.J. and Mohammed, S. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal Chem*, **84**, 9668-9673.

116. Tsiatsiani, L. and Heck, A.J. (2015) Proteomics beyond trypsin. *FEBS J*, **282**, 2612-2626.

117. Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature*, **537**, 347-355.

118. Munoz, J. and Heck, A.J.R. (2014) From the Human Genome to the Human Proteome. *Angew Chem Int Edit*, **53**, 10864-10866.

119. Wisniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nature Methods*, **6**, 359-U360.

120. Rappsilber, J., Mann, M. and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc*, **2**, 1896-1906.

121. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, **26**, 1367-1372.
122. Nesvizhskii, A.I. (2014) Proteogenomics: concepts, applications and computational strategies. *Nat Methods*, **11**, 1114-1125.
123. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C. and MacCoss, M.J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*, **26**, 966-968.
124. Olsen, J.V. and Mann, M. (2013) Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry. *Molecular & Cellular Proteomics*, **12**, 3444-3452.
125. Chambers, A.G., Percy, A.J., Yang, J. and Borchers, C.H. (2015) Multiple Reaction Monitoring Enables Precise Quantification of 97 Proteins in Dried Blood Spots. *Molecular & cellular proteomics : MCP*, **14**, 3094-3104.
126. Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D. and Mann, M. (2016) Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst*, **2**, 185-195.
127. Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., Thomas, J.K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N.A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L.D., Patil, A.H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S.K., Marimuthu, A., Sathe, G.J., Chavan, S., Datta, K.K., Subbannayya, Y., Sahu, A., Yelamanchi, S.D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K.R., Syed, N., Goel, R., Khan, A.A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A., Huang, T.C., Zhong, J., Wu, X., Shaw, P.G., Freed, D., Zahari, M.S., Mukherjee, K.K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C.J., Shankar, S.K., Satishchandra, P., Schroeder, J.T., Sirdeshmukh, R., Maitra, A., Leach, S.D., Drake, C.G., Halushka, M.K., Prasad, T.S., Hruban, R.H., Kerr, C.L., Bader, G.D., Iacobuzio-Donahue, C.A., Gowda, H. and Pandey, A. (2014) A draft map of the human proteome. *Nature*, **509**, 575-581.
128. Kingsmore, S.F. (2006) Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nat Rev Drug Discov*, **5**, 310-320.
129. Landegren, U., Vanelid, J., Hammond, M., Nong, R.Y., Wu, D., Ulleras, E. and Kamali-Moghaddam, M. (2012) Opportunities for sensitive plasma proteome analysis. *Anal Chem*, **84**, 1824-1830.
130. Gingras, A.C., Aebersold, R. and Raught, B. (2005) Advances in protein complex analysis using mass spectrometry. *J Physiol*, **563**, 11-21.
131. Anderson, N.L. and Anderson, N.G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Molecular & cellular proteomics : MCP*, **1**, 845-867.
132. Anderson, L. (2014) Six decades searching for meaning in the proteome. *J Proteomics*, **107**, 24-30.
133. Uhlen, M., Bandrowski, A., Carr, S., Edwards, A., Ellenberg, J., Lundberg, E., Rimm, D.L., Rodriguez, H., Hiltke, T., Snyder, M. and Yamamoto, T. (2016) A proposal for validation of antibodies. *Nat Methods*, **13**, 823-827.
134. Anderson, N.L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R.S., Conrads, T.P., Veenstra, T.D., Adkins, J.N., Pounds, J.G., Fagan, R. and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Molecular & cellular proteomics : MCP*, **3**, 311-326.
135. Bell, A.W., Deutsch, E.W., Au, C.E., Kearney, R.E., Beavis, R., Sechi, S., Nilsson, T., Bergeron, J.J. and Group, H.T.S.W. (2009) A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat Methods*, **6**, 423-430.
136. Branca, R.M., Orre, L.M., Johansson, H.J., Granholm, V., Huss, M., Perez-Bercoff, A., Forshed, J., Kall, L. and Lehtio, J. (2014) HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods*, **11**, 59-62.
137. Lesur, A. and Domon, B. (2015) Advances in high-resolution accurate mass spectrometry application to targeted proteomics. *Proteomics*, **15**, 880-890.
138. Todd, J., Freese, B., Lu, A., Held, D., Morey, J., Livingston, R. and Goix, P. (2007) Ultrasensitive flow-based immunoassays using single-molecule counting. *Clinical Chemistry*, **53**, 1990-1995.

139. Juncker, D., Bergeron, S., Laforte, V. and Li, H.Y. (2014) Cross-reactivity in antibody microarrays and multiplexed sandwich assays: shedding light on the dark side of multiplexing. *Curr Opin Chem Biol*, **18**, 29-37.

140. Holm, A., Wu, W.W. and Lund-Johansen, F. (2012) Antibody array analysis of labelled proteomes: how should we control specificity? *New Biotechnol*, **29**, 578-585.

141. Pla-Roca, M., Leulmi, R.F., Tourekhanova, S., Bergeron, S., Laforte, V., Moreau, E., Gosline, S.J., Bertos, N., Hallett, M., Park, M. and Juncker, D. (2012) Antibody colocalization microarray: a scalable technology for multiplex protein analysis in complex samples. *Molecular & cellular proteomics : MCP*, **11**, M111 011460.

142. Poetz, O., Henzler, T., Hartmann, M., Kazmaier, C., Templin, M.F., Herget, T. and Joos, T.O. (2010) Sequential multiplex analyte capturing for phosphoprotein profiling. *Molecular & cellular proteomics : MCP*, **9**, 2474-2481.

143. Ayoglu, B., Birgersson, E., Mezger, A., Nilsson, M., Uhlen, M., Nilsson, P. and Schwenk, J.M. (2016) Multiplexed protein profiling by sequential affinity capture. *Proteomics*, **16**, 1251-1256.

144. Boguski, M.S. and McIntosh, M.W. (2003) Biomedical informatics for proteomics. *Nature*, **422**, 233-237.

145. Marx, V. (2013) Targeted proteomics. *Nat Methods*, **10**, 19-22.

146. Crutchfield, C.A., Thomas, S.N., Sokoll, L.J. and Chan, D.W. (2016) Advances in mass spectrometry-based clinical biomarker discovery. *Clin Proteomics*, **13**, 1.

147. von Behring, E. and Kitasato, S. (1991) [The mechanism of diphtheria immunity and tetanus immunity in animals. 1890]. *Molecular immunology*, **28**, 1317, 1319-1320.

148. Porter, R.R. (1959) The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. *Biochem J*, **73**, 119-126.

149. Owen, J.A., Punt, J., Stranford, S.A., Jones, P.P. and Kuby, J. (2013) *Kuby immunology*. 7th ed. W.H. Freeman, New York.

150. Schroeder, H.W., Jr. and Cavacini, L. (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol*, **125**, S41-52.

151. Haraldsson, A., Kock-Jansen, M.J., Jaminon, M., van Eck-Arts, P.B., de Boo, T., Weemaes, C.M. and Bakkeren, J.A. (1991) Determination of kappa and lambda light chains in serum immunoglobulins G, A and M. *Ann Clin Biochem*, **28 (Pt 5)**, 461-466.

152. Matzinger, P. (2002) The danger model: a renewed sense of self. *Science*, **296**, 301-305.

153. Casadevall, A. and Janda, A. (2012) Immunoglobulin isotype influences affinity and specificity. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 12272-12273.

154. Janda, A. and Casadevall, A. (2010) Circular Dichroism reveals evidence of coupling between immunoglobulin constant and variable region secondary structure. *Molecular immunology*, **47**, 1421-1425.

155. Davies, D.R. and Cohen, G.H. (1996) Interactions of protein antigens with antibodies. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 7-12.

156. MacKenzie, C.R., Hirama, T., Deng, S.J., Bundle, D.R., Narang, S.A. and Young, N.M. (1996) Analysis by surface plasmon resonance of the influence of valence on the ligand binding affinity and kinetics of an anti-carbohydrate antibody. *J Biol Chem*, **271**, 1527-1533.

157. Vauquelin, G. and Charlton, S.J. (2013) Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands. *Br J Pharmacol*, **168**, 1771-1785.

158. Karush, F. (1970) Affinity and the immune response. *Annals of the New York Academy of Sciences*, **169**, 56-64.

159. Hornick, C.L. and Karush, F. (1972) Antibody Affinity .3. Role of Multivalence. *Immunochemistry*, **9**, 325-&.

160. Allen, T.M. (2002) Ligand-targeted therapeutics in anticancer therapy. *Nat Rev Cancer*, **2**, 750-763.

161. Rudnick, S.I. and Adams, G.P. (2009) Affinity and avidity in antibody-based tumor targeting. *Cancer Biother Radiopharm*, **24**, 155-161.

162. Klotz, I.M., Walker, F.M. and Pivan, R.B. (1946) The Binding of Organic Ions by Proteins. *J Am Chem Soc*, **68**, 1486-1490.
163. Liedberg, B., Nylander, C. and Lundstrom, I. (1983) Surface-Plasmon Resonance for Gas-Detection and Biosensing. *Sensor Actuator*, **4**, 299-304.
164. Friguet, B., Chaffotte, A.F., Djavadi-Ohaniance, L. and Goldberg, M.E. (1985) Measurements of the true affinity constant in solution of antigen-antibody complexes by enzyme-linked immunosorbent assay. *J Immunol Methods*, **77**, 305-319.
165. Jerabek-Willemsen, M., Andre, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P. and Breitsprecher, D. (2014) MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct*, **1077**, 101-113.
166. Reverberi, R. and Reverberi, L. (2007) Factors affecting the antigen-antibody reaction. *Blood Transfus*, **5**, 227-240.
167. Janeway, C. (2005) *Immunobiology : the immune system in health and disease*. 6th ed. Garland Science, New York.
168. Sundberg, E.J., Urrutia, M., Braden, B.C., Isern, J., Tsuchiya, D., Fields, B.A., Malchiodi, E.L., Tormo, J., Schwarz, F.P. and Mariuzza, R.A. (2000) Estimation of the hydrophobic effect in an antigen-antibody protein-protein interface. *Biochemistry*, **39**, 15375-15387.
169. Hozumi, N. and Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proceedings of the National Academy of Sciences of the United States of America*, **73**, 3628-3632.
170. Lefranc, M.-P., Protat, C., Jabado-Michaloud, J. and Ginestoux, C.
171. De Silva, N.S. and Klein, U. (2015) Dynamics of B cells in germinal centres. *Nature reviews. Immunology*, **15**, 137-148.
172. Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F. and Scharff, M.D. (2008) The biochemistry of somatic hypermutation. *Annu Rev Immunol*, **26**, 481-511.
173. Pasteur, L., Chamberland and Roux. (2002) Summary report of the experiments conducted at Pouilly-le-Fort, near Melun, on the anthrax vaccination, 1881. *Yale J Biol Med*, **75**, 59-62.
174. Smith, K.A. (2012) Louis pasteur, the father of immunology? *Front Immunol*, **3**, 68.
175. Charrin, A. and Roger, G.H. (1889) In Masson, G. (ed.), *Comptes rendus hebdomadaires de la société de biologie - 9e série.*, Paris, pp. 667.
176. Kantha, S.S. (1991) A centennial review; the 1890 tetanus antitoxin paper of von Behring and Kitasato and the related developments. *Keio J Med*, **40**, 35-39.
177. (1896) Report of the Lancet Special Commission on the Relative Strengths of Diphtheria Antitoxic Serums. *The Lancet*, **148**, 182-195.
178. Durham, H.E. (1896) On a Special Action of the Serum of Highly Immunised Animals, and Its Use for Diagnostic and other Purposes. *Proceedings of the Royal Society of London*, **59**, 224-226.
179. Grünbaum, A. (1896) Preliminary Note on the Use of the Agglutinative Action of Human Serum for the Diagnosis of Enteric Fever. *The Lancet*, **148**, 806-807.
180. Johnston, W. and McTaggart, D.D. (1896) Observations on the Serum Reaction in Typhoid Fever and Experimental Cholera by the Dried Blood Method. *Br Med J*, **2**, 1629-1631.
181. Widal, F. (1896) On the Sero-Diagnosis of Typhoid Fever. *The Lancet*, **148**, 1371-1372.
182. Delepine, S. (1896) On the "Sero-Diagnosis" of Typhoid Fever. *The Lancet*, **148**, 1587-1589.
183. Talmage, D.W. (1959) Immunological specificity, unique combinations of selected natural globulins provide an alternative to the classical concept. *Science*, **129**, 1643-1648.
184. Tiselius, A. and Kabat, E.A. (1939) An Electrophoretic Study of Immune Sera and Purified Antibody Preparations. *The Journal of experimental medicine*, **69**, 119-131.
185. Yalow, R.S. and Berson, S.A. (1970) Radioimmunoassay of gastrin. *Gastroenterology*, **58**, 1-14.
186. Sternberger, L.A., Hardy, P.H., Jr., Cuculis, J.J. and Meyer, H.G. (1970) The unlabeled antibody enzyme method of immunohistochemistry: preparation and properties of soluble antigen-antibody complex (horseradish peroxidase-antihorseradish peroxidase) and its use in identification of spirochetes. *J Histochem Cytochem*, **18**, 315-333.
187. Talmage, D.W. (1957) Diversity of antibodies. *J Cell Physiol Suppl*, **50**, 229-246.

188. Landsteiner, K. and van der Scheer, J. (1936) On Cross Reactions of Immune Sera to Azoproteins. *The Journal of experimental medicine*, **63**, 325-339.
189. Heidelberger, M., Kabat, E.A. and Mayer, M. (1942) A Further Study of the Cross Reaction between the Specific Polysaccharides of Types Iii and Viii Pneumococci in Horse Antisera. *The Journal of experimental medicine*, **75**, 35-47.
190. Cuatrecasas, P. (1969) Insulin-Sepharose: Immunoreactivity and use in the purification of antibody. *Biochemical and Biophysical Research Communications*, **35**, 531-537.
191. Forsgren, A. and Sjoquist, J. (1966) "Protein A" from S. aureus. I. Pseudo-immune reaction with human gamma-globulin. *Journal of immunology*, **97**, 822-827.
192. Sjobring, U., Bjorck, L. and Kastern, W. (1991) Streptococcal protein G. Gene structure and protein binding properties. *J Biol Chem*, **266**, 399-405.
193. Köhler, G. and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, **256**, 495-497.
194. Askonas, B.A., Williamson, A.R. and Wright, B.E. (1970) Selection of a single antibody-forming cell clone and its propagation in syngeneic mice. *Proceedings of the National Academy of Sciences of the United States of America*, **67**, 1398-1403.
195. Playfair, J.H., Papermaster, B.W. and Cole, L.J. (1965) Focal antibody production by transferred spleen cells in irradiated mice. *Science*, **149**, 998-1000.
196. Lipman, N.S., Jackson, L.R., Trudel, L.J. and Weis-Garcia, F. (2005) Monoclonal versus polyclonal antibodies: distinguishing characteristics, applications, and information resources. *ILAR J*, **46**, 258-268.
197. George, A.J.T. and Urch, C.E. (2000) *Diagnostic and therapeutic antibodies*. Humana Press, Totowa, N.J.
198. Sevier, E.D., David, G.S., Martinis, J., Desmond, W.J., Bartholomew, R.M. and Wang, R. (1981) Monoclonal antibodies in clinical immunology. *Clin Chem*, **27**, 1797-1806.
199. James, L.C. and Tawfik, D.S. (2003) The specificity of cross-reactivity: Promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Sci*, **12**, 2183-2193.
200. Van Regenmortel, M.H. (2016) Structure-Based Reverse Vaccinology Failed in the Case of HIV Because it Disregarded Accepted Immunological Theory. *Int J Mol Sci*, **17**.
201. Van Regenmortel, M.H.V. (2014) Specificity, polyspecificity, and heterospecificity of antibody-antigen recognition. *Journal of Molecular Recognition*, **27**, 627-639.
202. Hjelm, B., Forsstrom, B., Igel, U., Johannesson, H., Stadler, C., Lundberg, E., Ponten, F., Sjoberg, A., Rockberg, J., Schwenk, J.M., Nilsson, P., Johansson, C. and Uhlen, M. (2011) Generation of monospecific antibodies based on affinity capture of polyclonal antibodies. *Protein Sci*, **20**, 1824-1835.
203. Pope, M.E., Soste, M.V., Eyford, B.A., Anderson, N.L. and Pearson, T.W. (2009) Anti-peptide antibody screening: selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique. *J Immunol Methods*, **341**, 86-96.
204. Holliger, P. and Hudson, P.J. (2005) Engineered antibody fragments and the rise of single domains. *Nature biotechnology*, **23**, 1126-1136.
205. Helma, J., Cardoso, M.C., Muyldermans, S. and Leonhardt, H. (2015) Nanobodies and recombinant binders in cell biology. *J Cell Biol*, **209**, 633-644.
206. Bordeaux, J., Welsh, A., Agarwal, S., Killiam, E., Baquero, M., Hanna, J., Anagnostou, V. and Rimm, D. (2010) Antibody validation. *Biotechniques*, **48**, 197-209.
207. Baker, M. (2015) Reproducibility crisis: Blame it on the antibodies. *Nature*, **521**, 274-276.
208. Colwill, K., Renewable Protein Binder Working, G. and Graslund, S. (2011) A roadmap to generate renewable protein binders to the human proteome. *Nat Methods*, **8**, 551-558.
209. Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R.A., Wilton, J. and Johnson, K.S. (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature biotechnology*, **14**, 309-314.
210. Weller, M.G. (2016) Quality Issues of Research Antibodies. *Anal Chem Insights*, **11**, 21-27.

211. Raghunathan, G., Smart, J., Williams, J. and Almagro, J.C. (2012) Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens. *J Mol Recognit*, **25**, 103-113.
212. Kringelum, J.V., Nielsen, M., Padkjaer, S.B. and Lund, O. (2013) Structural analysis of B-cell epitopes in antibody:protein complexes. *Molecular immunology*, **53**, 24-34.
213. Ofran, Y., Schlessinger, A. and Rost, B. (2008) Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. *Journal of immunology*, **181**, 6230-6235.
214. Ramaraj, T., Angel, T., Dratz, E.A., Jesaitis, A.J. and Mumey, B. (2012) Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures. *Biochim Biophys Acta*, **1824**, 520-532.
215. Kunik, V. and Ofran, Y. (2013) The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Eng Des Sel*, **26**, 599-609.
216. Rockberg, J. and Uhlen, M. (2009) Prediction of antibody response using recombinant human protein fragments as antigen. *Protein Sci*, **18**, 2346-2355.
217. Rubinstein, N.D., Mayrose, I., Halperin, D., Yekutieli, D., Gershoni, J.M. and Pupko, T. (2008) Computational characterization of B-cell epitopes. *Molecular immunology*, **45**, 3477-3489.
218. Viart, B., Dias-Lopes, C., Kozlova, E., Oliveira, C.F.B., Nguyen, C., Neshich, G., Chavez-Olortegui, C., Molina, F. and Felicori, L.F. (2016) EPI-peptide designer: a tool for designing peptide ligand libraries based on epitope-paratope interactions. *Bioinformatics*, **32**, 1462-1470.
219. Collis, A.V.J., Brouwer, A.P. and Martin, A.C.R. (2003) Analysis of the antigen combining site: Correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *Journal of Molecular Biology*, **325**, 337-354.
220. Stave, J.W. and Lindpaintner, K. (2013) Antibody and antigen contact residues define epitope and paratope size and structure. *Journal of immunology*, **191**, 1428-1435.
221. Regenmortel, M.H.V.V. (1998) Mimotopes, continuous paratopes and hydropathic complementarity: novel approximations in the description of immunochemical specificity. *Journal of Dispersion Science and Technology*, **19**, 1199-1219.
222. Geysen, H.M., Mason, T.J. and Rodda, S.J. (1988) Cognitive features of continuous antigenic determinants. *J Mol Recognit*, **1**, 32-41.
223. Williams, W.V., Guy, H.R., Rubin, D.H., Robey, F., Myers, J.N., Kieberemmons, T., Weiner, D.B. and Greene, M.I. (1988) Sequences of the Cell-Attachment Sites of Reovirus Type-3 and Its Anti-Idiotypic Antireceptor Antibody - Modeling of Their 3-Dimensional Structures. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 6488-6492.
224. Laune, D., Molina, F., Ferrieres, G., Mani, J.C., Cohen, P., Simon, D., Bernardi, T., Piechaczyk, M., Pau, B. and Granier, C. (1997) Systematic exploration of the antigen binding-activity of synthetic peptides isolated from the variable regions of immunoglobulins. *Journal of Biological Chemistry*, **272**, 30937-30944.
225. Laune, D., Molina, F., Ferrieres, G., Villard, S., Bes, C., Rieunier, F., Chardes, T. and Granier, C. (2002) Application of the Spot method to the identification of peptides and amino acids from the antibody paratope that contribute to antigen binding. *Journal of Immunological Methods*, **267**, 53-70.
226. Dimitrov, J.D., Pashov, A.D. and Vassilev, T.L. (2012) ANTIBODY POLYSPECIFICITY: What Does It Matter? *Adv Exp Med Biol*, **750**, 213-226.
227. Keitel, T., Kramer, A., Wessner, H., Scholz, C., SchneiderMergener, J. and Hohne, W. (1997) Crystallographic analysis of anti-p24 (HIV-1) monoclonal antibody cross-reactivity and polyspecificity. *Cell*, **91**, 811-820.
228. Van Regenmortel, M.H.V. (2012) Requirements for empirical immunogenicity trials, rather than structure-based design, for developing an effective HIV vaccine. *Arch Virol*, **157**, 1-20.
229. Cohn, M. (2005) Degeneracy, mimicry and crossreactivity in immune recognition. *Molecular immunology*, **42**, 651-655.
230. Oldstone, M.B.A. (1998) Molecular mimicry and immune-mediated diseases. *Faseb J*, **12**, 1255-1265.

231. Albert, L.J. and Inman, R.D. (1999) Mechanisms of disease: Molecular mimicry and autoimmunity. *New Engl J Med*, **341**, 2068-2074.

232. Cusick, M.F., Libbey, J.E. and Fujinami, R.S. (2012) Molecular Mimicry as a Mechanism of Autoimmune Disease. *Clinical reviews in allergy & immunology*, **42**, 102-111.

233. Robinson, W.H. (2015) Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. *Nat Rev Rheumatol*, **11**, 171-182.

234. Dorner, T., Jacobi, A.M. and Lipsky, P.E. (2009) B cells in autoimmunity. *Arthritis Res Ther*, **11**, 247.

235. von Boehmer, H. and Melchers, F. (2010) Checkpoints in lymphocyte development and autoimmune disease. *Nat Immunol*, **11**, 14-20.

236. Rawlings, D.J., Metzler, G., Wray-Dutra, M. and Jackson, S.W. (2017) Altered B cell signalling in autoimmunity. *Nature reviews. Immunology*, **17**, 421-436.

237. Hogquist, K.A., Baldwin, T.A. and Jameson, S.C. (2005) Central tolerance: learning self-control in the thymus. *Nature reviews. Immunology*, **5**, 772-782.

238. Hogquist, K.A. and Jameson, S.C. (2014) The self-obsession of T cells: how TCR signaling thresholds affect fate 'decisions' and effector function. *Nat Immunol*, **15**, 815-823.

239. Theofilopoulos, A.N., Kono, D.H. and Baccala, R. (2017) The multiple pathways to autoimmunity. *Nat Immunol*, **18**, 716-724.

240. Mignot, E., Lin, L., Rogers, W., Honda, Y., Qiu, X., Lin, X., Okun, M., Hohjoh, H., Miki, T., Hsu, S., Leffell, M., Grumet, F., Fernandez-Vina, M., Honda, M. and Risch, N. (2001) Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. *Am J Hum Genet*, **68**, 686-699.

241. Graham, R.R., Kozyrev, S.V., Baechler, E.C., Reddy, M.V., Plenge, R.M., Bauer, J.W., Ortmann, W.A., Koeuth, T., Gonzalez Escribano, M.F., Argentine, Spanish Collaborative, G., Pons-Estel, B., Petri, M., Daly, M., Gregersen, P.K., Martin, J., Altshuler, D., Behrens, T.W. and Alarcon-Riquelme, M.E. (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nature genetics*, **38**, 550-555.

242. Cho, J.H. and Gregersen, P.K. (2011) Genomics and the multifactorial nature of human autoimmune disease. *The New England journal of medicine*, **365**, 1612-1623.

243. Salvetti, M., Ristori, G., Bomprezzi, R., Pozzilli, P. and Leslie, R.D. (2000) Twins: mirrors of the immune system. *Immunol Today*, **21**, 342-347.

244. Heliovaara, M., Aho, K., Aromaa, A., Knekt, P. and Reunanen, A. (1993) Smoking and risk of rheumatoid arthritis. *J Rheumatol*, **20**, 1830-1835.

245. Silman, A.J., Newman, J. and MacGregor, A.J. (1996) Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. *Arthritis and rheumatism*, **39**, 732-735.

246. Costenbader, K.H., Kim, D.J., Peerzada, J., Lockman, S., Nobles-Knight, D., Petri, M. and Karlson, E.W. (2004) Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis. *Arthritis and rheumatism*, **50**, 849-857.

247. Adorini, L. and Penna, G. (2008) Control of autoimmune diseases by the vitamin D endocrine system. *Nat Clin Pract Rheumatol*, **4**, 404-412.

248. Gleicher, N. and Barad, D.H. (2007) Gender as risk factor for autoimmune diseases. *J Autoimmun*, **28**, 1-6.

249. Zandman-Goddard, G., Peeva, E. and Shoenfeld, Y. (2007) Gender and autoimmunity. *Autoimmunity reviews*, **6**, 366-372.

250. Markle, J.G., Frank, D.N., Mortin-Toth, S., Robertson, C.E., Feazel, L.M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K.D., Macpherson, A.J. and Danska, J.S. (2013) Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science*, **339**, 1084-1088.

251. Kamada, N., Seo, S.U., Chen, G.Y. and Nunez, G. (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nature reviews. Immunology*, **13**, 321-335.

252. Shoenfeld, Y., Agmon-Levin, N. and Tomljevic, L. (2015) *Vaccines and autoimmunity*. Wiley Blackwell, Hoboken, New Jersey.

253. Ruff, W.E. and Kriegel, M.A. (2015) Autoimmune host-microbiota interactions at barrier sites and beyond. *Trends Mol Med*, **21**, 233-244.

254. Shlomchik, M.J., Craft, J.E. and Mamula, M.J. (2001) From T to B and back again: positive feedback in systemic autoimmune disease. *Nature reviews. Immunology*, **1**, 147-153.

255. Vanderlugt, C.L. and Miller, S.D. (2002) Epitope spreading in immune-mediated diseases: implications for immunotherapy. *Nature reviews. Immunology*, **2**, 85-95.

256. Kurosaki, T., Kometani, K. and Ise, W. (2015) Memory B cells. *Nature reviews. Immunology*, **15**, 149-159.

257. Olson, J.K., Croxford, J.L., Calenoff, M.A., Dal Canto, M.C. and Miller, S.D. (2001) A virus-induced molecular mimicry model of multiple sclerosis. *J Clin Invest*, **108**, 311-318.

258. Horwitz, M.S., Bradley, L.M., Harbertson, J., Krahl, T., Lee, J. and Sarvetnick, N. (1998) Diabetes induced by Coxsackie virus: initiation by bystander damage and not molecular mimicry. *Nature medicine*, **4**, 781-785.

259. Mor, F. and Cohen, I.R. (1995) Pathogenicity of T cells responsive to diverse cryptic epitopes of myelin basic protein in the Lewis rat. *Journal of immunology*, **155**, 3693-3699.

260. Lanzavecchia, A. (1995) How can cryptic epitopes trigger autoimmunity? *The Journal of experimental medicine*, **181**, 1945-1948.

261. McMahon, E.J., Bailey, S.L., Castenada, C.V., Waldner, H. and Miller, S.D. (2005) Epitope spreading initiates in the CNS in two mouse models of multiple sclerosis. *Nature medicine*, **11**, 335-339.

262. Cornaby, C., Gibbons, L., Mayhew, V., Sloan, C.S., Welling, A. and Poole, B.D. (2015) B cell epitope spreading: mechanisms and contribution to autoimmune diseases. *Immunol Lett*, **163**, 56-68.

263. Witebsky, E., Rose, N.R., Terplan, K., Paine, J.R. and Egan, R.W. (1957) Chronic thyroiditis and autoimmunization. *Journal of the American Medical Association*, **164**, 1439-1447.

264. Suurmond, J. and Diamond, B. (2015) Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity. *J Clin Invest*, **125**, 2194-2202.

265. Di Lorenzo, T.P., Peakman, M. and Roep, B.O. (2007) Translational mini-review series on type 1 diabetes: Systematic analysis of T cell epitopes in autoimmune diabetes. *Clin Exp Immunol*, **148**, 1-16.

266. Tan, E.M. (1989) Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv Immunol*, **44**, 93-151.

267. Fernandez Madrid, F. (2005) Autoantibodies in breast cancer sera: candidate biomarkers and reporters of tumorigenesis. *Cancer Lett*, **230**, 187-198.

268. Ludvigsson, J. (2009) Therapy with GAD in diabetes. *Diabetes Metab Res Rev*, **25**, 307-315.

269. Thurau, S.R., Diedrichs-Mohring, M., Fricke, H., Burchardi, C. and Wildner, G. (1999) Oral tolerance with an HLA-peptide mimicking retinal autoantigen as a treatment of autoimmune uveitis. *Immunol Lett*, **68**, 205-212.

270. Henry, R.A., Kendall, P.L. and Thomas, J.W. (2012) Autoantigen-specific B-cell depletion overcomes failed immune tolerance in type 1 diabetes. *Diabetes*, **61**, 2037-2044.

271. Willemze, A., Trouw, L.A., Toes, R.E. and Huizinga, T.W. (2012) The influence of ACPA status and characteristics on the course of RA. *Nat Rev Rheumatol*, **8**, 144-152.

272. Vanderlugt, C.L., Neville, K.L., Nikcevich, K.M., Eagar, T.N., Bluestone, J.A. and Miller, S.D. (2000) Pathologic role and temporal appearance of newly emerging autoepitopes in relapsing experimental autoimmune encephalomyelitis. *Journal of immunology*, **164**, 670-678.

273. Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A. and Wilson, I.A. (1992) Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science*, **257**, 919-927.

274. Ponomarenko, J.V. and Bourne, P.E. (2007) Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *Bmc Struct Biol*, **7**.

275. Potocnakova, L., Bhide, M. and Pulzova, L.B. (2016) An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *J Immunol Res*, **2016**, 6760830.

276. Hino, T., Iwata, S. and Murata, T. (2013) Generation of functional antibodies for mammalian membrane protein crystallography. *Curr Opin Struct Biol*, **23**, 563-568.

277. Benjamin, D.C. and Perdue, S.S. (1996) Site-Directed Mutagenesis in Epitope Mapping. *Methods*, **9**, 508-515.
278. Smith-Gill, S.J. (1994) Protein epitopes: functional vs. structural definitions. *Res Immunol*, **145**, 67-70.
279. Dougan, D.A., Malby, R.L., Gruen, L.C., Kortt, A.A. and Hudson, P.J. (1998) Effects of substitutions in the binding surface of an antibody on antigen affinity. *Protein Eng*, **11**, 65-74.
280. Hoylaerts, M.F. and Millan, J.L. (1991) Site-directed mutagenesis and epitope-mapped monoclonal antibodies define a catalytically important conformational difference between human placental and germ cell alkaline phosphatase. *Eur J Biochem*, **202**, 605-616.
281. Cunningham, B.C. and Wells, J.A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081-1085.
282. Batista, F.D. and Neuberger, M.S. (1998) Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. *Immunity*, **8**, 751-759.
283. Opuni, K.F., Al-Majdoub, M., Yefremova, Y., El-Kased, R.F., Koy, C. and Glocker, M.O. (2016) Mass spectrometric epitope mapping. *Mass Spectrom Rev.*
284. Hager-Braun, C. and Tomer, K.B. (2005) Determination of protein-derived epitopes by mass spectrometry. *Expert Rev Proteomics*, **2**, 745-756.
285. Ladner, R.C. (2007) Mapping the epitopes of antibodies. *Biotechnol Genet Eng Rev*, **24**, 1-30.
286. Yang, D., Frego, L., Lasaro, M., Truncali, K., Kroe-Barrett, R. and Singh, S. (2016) Efficient Qualitative and Quantitative Determination of Antigen-induced Immune Responses. *J Biol Chem*, **291**, 16361-16374.
287. Stawikowski, M. and Fields, G.B. (2012) Introduction to peptide synthesis. *Curr Protoc Protein Sci*, **Chapter 18**, Unit 18 11.
288. Reineke, U. and Schutkowski, M. (2009) *Epitope mapping protocols*. 2nd ed. Humana Press, New York.
289. Atassi, M.Z. and Pai, R.C. (1975) Immunochemistry of sperm-whale myoglobin--XXII. Accurate delineation of the single reactive region in sequence 103-120 by immunochemical studies of synthetic peptides: the complete antigenic structure of the protein. *Immunochemistry*, **12**, 735-740.
290. Geysen, H.M., Meloen, R.H. and Barteling, S.J. (1984) Use of Peptide-Synthesis to Probe Viral-Antigens for Epitopes to a Resolution of a Single Amino-Acid. *P Natl Acad Sci-Biol*, **81**, 3998-4002.
291. Heiko, A. and Carsten, G. (2009) Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. *Current Proteomics*, **6**, 1-12.
292. Ahmad, T.A., Eweida, A.E. and Sheweita, S.A. (2016) B-cell epitope mapping for the design of vaccines and effective diagnostics. *Trials in Vaccinology*, **5**, 71-83.
293. Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T. and Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science*, **251**, 767-773.
294. Legutki, J.B., Zhao, Z.G., Greving, M., Woodbury, N., Johnston, S.A. and Stafford, P. (2014) Scalable high-density peptide arrays for comprehensive health monitoring. *Nature communications*, **5**, 4785.
295. Pellois, J.P., Zhou, X., Srivannavit, O., Zhou, T., Gulari, E. and Gao, X. (2002) Individually addressable parallel peptide synthesis on microchips. *Nature biotechnology*, **20**, 922-926.
296. Carmona, S.J., Nielsen, M., Schafer-Nielsen, C., Mucci, J., Altcheh, J., Balouz, V., Tekiel, V., Frasch, A.C., Campetella, O., Buscaglia, C.A. and Aguero, F. (2015) Towards High-throughput Immunomics for Infectious Diseases: Use of Next-generation Peptide Microarrays for Rapid Discovery and Mapping of Antigenic Determinants. *Molecular & cellular proteomics : MCP*, **14**, 1871-1884.
297. O'Donnell, B., Maurer, A., Papandreou-Suppappola, A. and Stafford, P. (2015) Time-Frequency Analysis of Peptide Microarray Data: Application to Brain Cancer Immunosignatures. *Cancer Inform*, **14**, 219-233.
298. Luzzago, A., Felici, F., Tramontano, A., Pessi, A. and Cortese, R. (1993) Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage library of constrained peptides. *Gene*, **128**, 51-57.

299. Prezzi, C., Nuzzo, M., Meola, A., Delmastro, P., Galfre, G., Cortese, R., Nicosia, A. and Monaci, P. (1996) Selection of antigenic and immunogenic mimics of hepatitis C virus using sera from patients. *Journal of immunology*, **156**, 4504-4513.

300. Folgori, A., Tafi, R., Meola, A., Felici, F., Galfre, G., Cortese, R., Monaci, P. and Nicosia, A. (1994) A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *EMBO J*, **13**, 2236-2243.

301. Richer, J., Johnston, S.A. and Stafford, P. (2015) Epitope identification from fixed-complexity random-sequence peptide microarrays. *Molecular & cellular proteomics : MCP*, **14**, 136-147.

302. Hueber, W. and Robinson, W.H. (2006) Proteomic biomarkers for autoimmune disease. *Proteomics*, **6**, 4100-4105.

303. Zhu, H. and Snyder, M. (2001) Protein arrays and microarrays. *Curr Opin Chem Biol*, **5**, 40-45.

304. Sutandy, F.X., Qian, J., Chen, C.S. and Zhu, H. (2013) Overview of protein microarrays. *Curr Protoc Protein Sci*, **Chapter 27**, Unit 27 21.

305. Structural Genomics, C., China Structural Genomics, C., Northeast Structural Genomics, C., Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H.W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.H., Rao, Z., Shi, Y., Terwilliger, T.C., Kim, C.Y., Hung, L.W., Waldo, G.S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J.L., Stevens, R.C., Lesley, S.A., Wilson, I.A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M.I., Eschenfeldt, W.H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S.K., Emtage, J.S., Sauder, J.M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S.C., Bonanno, J.B., Fiser, A., Swaminathan, S., Studier, F.W., Chance, M.R., Sali, A., Acton, T.B., Xiao, R., Zhao, L., Ma, L.C., Hunt, J.F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C.K., Wang, D., Wang, H., Jiang, M., Montelione, G.T., Stuart, D.I., Owens, R.J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K. and Gunsalus, K.C. (2008) Protein production and purification. *Nat Methods*, **5**, 135-146.

306. Tegel, H., Steen, J., Konrad, A., Nikdin, H., Pettersson, K., Stenvall, M., Tourle, S., Wrethagen, U., Xu, L., Yderland, L., Uhlen, M., Hober, S. and Ottosson, J. (2009) High-throughput protein production--lessons from scaling up from 10 to 288 recombinant proteins per week. *Biotechnology journal*, **4**, 51-57.

307. He, M., Stoevesandt, O., Palmer, E.A., Khan, F., Ericsson, O. and Taussig, M.J. (2008) Printing protein arrays from DNA arrays. *Nat Methods*, **5**, 175-177.

308. Ramachandran, N., Raphael, J.V., Hainsworth, E., Demirkan, G., Fuentes, M.G., Rolfs, A., Hu, Y. and LaBaer, J. (2008) Next-generation high-density self-assembling functional protein arrays. *Nat Methods*, **5**, 535-538.

309. Meyer, S., Woodward, M., Hertel, C., Vlaicu, P., Haque, Y., Karner, J., Macagno, A., Onuoha, S.C., Fishman, D., Peterson, H., Metskula, K., Uibo, R., Jantti, K., Hokynar, K., Wolff, A.S., collaborative, A.p., Krohn, K., Ranki, A., Peterson, P., Kisand, K. and Hayday, A. (2016) AIRE-Deficient Patients Harbor Unique High-Affinity Disease-Ameliorating Autoantibodies. *Cell*, **166**, 582-595.

310. Sjoberg, R., Mattsson, C., Andersson, E., Hellstrom, C., Uhlen, M., Schwenk, J.M., Ayoglu, B. and Nilsson, P. (2016) Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. *N Biotechnol*, **33**, 582-592.

311. Duarte, J.G. and Blackburn, J.M. (2017) Advances in the development of human protein microarrays. *Expert Rev Proteomic*, **14**, 627-641.

312. Hu, C.J., Pan, J.B., Song, G., Wen, X.T., Wu, Z.Y., Chen, S., Mo, W.X., Zhang, F.C., Qian, J., Zhu, H. and Li, Y.Z. (2017) Identification of Novel Biomarkers for Behcet Disease Diagnosis Using Human Proteome Microarray Approach. *Molecular & cellular proteomics : MCP*, **16**, 147-156.

313. van der Meulen, P.M., Barendregt, A.M., Cuadrado, E., Magro-Checa, C., Steup-Beekman, G.M., Schonenberg-Meinema, D., Van den Berg, J.M., Li, Q.Z., Baars, P.A., Wouters, D., Voskuyl, A.E., Ten Berge, I., Huizinga, T.W.J. and Kuijpers, T.W. (2017) Protein array autoantibody profiles to determine diagnostic markers for neuropsychiatric systemic lupus erythematosus. *Rheumatology (Oxford)*.

314. Auger, I., Balandraud, N., Rak, J., Lambert, N., Martin, M. and Roudier, J. (2009) New autoantigens in rheumatoid arthritis (RA): screening 8268 protein arrays with sera from patients with RA. *Ann Rheum Dis*, **68**, 591-594.

315. Landegren, N., Sharon, D., Shum, A.K., Khan, I.S., Fasano, K.J., Hallgren, A., Kampf, C., Freyhult, E., Ardesjo-Lundgren, B., Alimohammadi, M., Rathsman, S., Ludvigsson, J.F., Lundh, D., Motrich, R., Rivero, V., Fong, L., Giwercman, A., Gustafsson, J., Perheentupa, J., Husebye, E.S., Anderson, M.S., Snyder, M. and Kampe, O. (2015) Transglutaminase 4 as a prostate autoantigen in male subfertility. *Science translational medicine*, **7**, 292ra101.

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

317. Hoogenboom, H.R. (2005) Selecting and screening recombinant antibody libraries. *Nature biotechnology*, **23**, 1105-1116.

318. Löfblom, J. (2011) Bacterial display in combinatorial protein engineering. *Biotechnology journal*, **6**, 1115-1129.

319. Larman, H.B., Zhao, Z., Laserson, U., Li, M.Z., Ciccia, A., Gakidis, M.A., Church, G.M., Kesari, S., Leproust, E.M., Solimini, N.L. and Elledge, S.J. (2011) Autoantigen discovery with a synthetic human peptidome. *Nature biotechnology*, **29**, 535-541.

320. Xu, G.J., Kula, T., Xu, Q., Li, M.Z., Vernon, S.D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R.T., O'Connor, K.C., Walker, B., Larman, H.B. and Elledge, S.J. (2015) Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. *Science*, **348**, aaa0698.

321. Fersht, A.R. (2008) From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat Rev Mol Cell Biol*, **9**, 650-654.

322. Chou, P.Y. and Fasman, G.D. (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry*, **13**, 211-222.

323. Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 3824-3828.

324. Parker, J.M., Guo, D. and Hodges, R.S. (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, **25**, 5425-5432.

325. Karplus, P.A. and Schulz, G.E. (1985) Prediction of chain flexibility in proteins. *Naturwissenschaften*, **72**, 212-213.

326. Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol*, **55**, 836-839.

327. Welling, G.W., Weijer, W.J., van der Zee, R. and Welling-Wester, S. (1985) Prediction of sequential antigenic regions in proteins. *FEBS Lett*, **188**, 215-218.

328. Ponomarenko, J.V. and Van Regenmortel, M.H.V. (2009) In Gu, J. and Bourne, P. E. (eds.), *Structural bioinformatics*. 2nd ed. Wiley-Blackwell, Hoboken, N.J., pp. xxvi, 1035 p., 1032 p. of plates.

329. Blythe, M.J. and Flower, D.R. (2005) Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci*, **14**, 246-248.

330. Saha, S. and Raghava, G.P. (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, **65**, 40-48.

331. El-Manzalawy, Y., Dobbs, D. and Honavar, V. (2008) Predicting linear B-cell epitopes using string kernels. *J Mol Recognit*, **21**, 243-255.

332. Larsen, J.E., Lund, O. and Nielsen, M. (2006) Improved method for predicting linear B-cell epitopes. *Immunome Res*, **2**, 2.

333. Fleri, W., Paul, S., Dhanda, S.K., Mahajan, S., Xu, X., Peters, B. and Sette, A. (2017) The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Front Immunol*, **8**, 278.

334. Jespersen, M.C., Peters, B., Nielsen, M. and Marcatili, P. (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.*

335. Haste Andersen, P., Nielsen, M. and Lund, O. (2006) Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci*, **15**, 2558-2567.

336. Kringelum, J.V., Lundegaard, C., Lund, O. and Nielsen, M. (2012) Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *Plos Comput Biol*, **8**, e1002829.

337. Kulkarni-Kale, U., Bhosle, S. and Kolaskar, A.S. (2005) CEP: a conformational epitope prediction server. *Nucleic Acids Res*, **33**, W168-171.

338. Kornum, B.R., Faraco, J. and Mignot, E. (2011) Narcolepsy with hypocretin/orexin deficiency, infections and autoimmunity of the brain. *Curr Opin Neurobiol*, **21**, 897-903.

339. Tafti, M., Hor, H., Dauvilliers, Y., Lammers, G.J., Overeem, S., Mayer, G., Javidi, S., Iranzo, A., Santamaria, J., Peraita-Adrados, R., Vicario, J.L., Arnulf, I., Plazzi, G., Bayard, S., Poli, F., Pizza, F., Geisler, P., Wierzbicka, A., Bassetti, C.L., Mathis, J., Lecendreux, M., Donjacour, C.E., van der Heide, A., Heinzer, R., Haba-Rubio, J., Feketeova, E., Hogl, B., Frauscher, B., Beneto, A., Khatami, R., Canellas, F., Pfister, C., Scholz, S., Billiard, M., Baumann, C.R., Ercilla, G., Verduijn, W., Claas, F.H., Dubois, V., Nowak, J., Eberhard, H.P., Pradervand, S., Hor, C.N., Testi, M., Tiercy, J.M. and Kutalik, Z. (2014) DQB1 locus alone explains most of the risk and protection in narcolepsy with cataplexy in Europe. *Sleep*, **37**, 19-25.

340. Ahmed, S.S., Schur, P.H., MacDonald, N.E. and Steinman, L. (2014) Narcolepsy, 2009 A(H1N1) pandemic influenza, and pandemic influenza vaccinations: what is known and unknown about the neurological disorder, the role for autoimmunity, and vaccine adjuvants. *J Autoimmun*, **50**, 1-11.

341. Larman, H.B., Salajegheh, M., Nazareno, R., Lam, T., Sauld, J., Steen, H., Kong, S.W., Pinkus, J.L., Amato, A.A., Elledge, S.J. and Greenberg, S.A. (2013) Cytosolic 5'-nucleotidase 1A autoimmunity in sporadic inclusion body myositis. *Annals of neurology*, **73**, 408-418.

342. Lehmann-Facius, H. (1937) Über die Liquordiagnose der Schizophrenien. *Klinische Wochenschrift*, **16**, 1646-1648.

343. Knight, J.G., Menkes, D.B., Highton, J. and Adams, D.D. (2007) Rationale for a trial of immunosuppressive therapy in acute schizophrenia. *Mol Psychiatry*, **12**, 424-431.

344. Ellul, P., Groc, L., Tamouza, R. and Leboyer, M. (2017) The Clinical Challenge of Autoimmune Psychosis: Learning from Anti-NMDA Receptor Autoantibodies. *Front Psychiatry*, **8**, 54.

345. Ezeoke, A., Mellor, A., Buckley, P. and Miller, B. (2013) A systematic, quantitative review of blood autoantibodies in schizophrenia. *Schizophr Res*, **150**, 245-251.

346. Steinman, L. (1996) Multiple sclerosis: a coordinated immunological attack against myelin in the central nervous system. *Cell*, **85**, 299-302.

347. Wucherpfennig, K.W. and Strominger, J.L. (1995) Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell*, **80**, 695-705.

348. Pruitt, K.D., Harrow, J., Harte, R.A., Wallin, C., Diekhans, M., Maglott, D.R., Searle, S., Farrell, C.M., Loveland, J.E., Ruef, B.J., Hart, E., Suner, M.M., Landrum, M.J., Aken, B., Ayling, S., Baertsch, R., Fernandez-Banet, J., Cherry, J.L., Curwen, V., Dicuccio, M., Kellis, M., Lee, J., Lin, M.F., Schuster, M., Shkeda, A., Amid, C., Brown, G., Dukhanina, O., Frankish, A., Hart, J., Maidak, B.L., Mudge, J., Murphy, M.R., Murphy, T., Rajan, J., Rajput, B., Riddick, L.D., Snow, C., Steward, C., Webb, D., Weber, J.A., Wilming, L., Wu, W., Birney, E., Haussler, D., Hubbard, T., Ostell, J., Durbin, R. and Lipman, D. (2009) The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome research*, **19**, 1316-1323.

349. .

350. Petter, M., Haeggstrom, M., Khattab, A., Fernandez, V., Klinkert, M.Q. and Wahlgren, M. (2007) Variant proteins of the Plasmodium falciparum RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol Biochem Parasitol*, **156**, 51-61.

351. Carlson, J., Nash, G.B., Gabutti, V., al-Yaman, F. and Wahlgren, M. (1994) Natural protection against severe Plasmodium falciparum malaria due to impaired rosette formation. *Blood*, **84**, 3909-3914.

352. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.

353. Haggmark, A., Hamsten, C., Wiklundh, E., Lindskog, C., Mattsson, C., Andersson, E., Lundberg, I.E., Gronlund, H., Schwenk, J.M., Eklund, A., Grunewald, J. and Nilsson, P. (2015) Proteomic profiling reveals autoimmune targets in sarcoidosis. *American journal of respiratory and critical care medicine*, **191**, 574-583.

354. Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., Cheng, Y., Clark, M.J., Im, H., Habegger, L., Balasubramanian, S., O'Huallachain, M., Dudley, J.T., Hillenmeyer, S., Haraksingh, R., Sharon, D., Euskirchen, G., Lacroute, P., Bettinger, K., Boyle, A.P., Kasowski, M., Grubert, F., Seki, S., Garcia, M., Whirl-Carrillo, M., Gallardo, M., Blasco, M.A., Greenberg, P.L., Snyder, P., Klein, T.E., Altman, R.B., Butte, A.J., Ashley, E.A., Gerstein, M., Nadeau, K.C., Tang, H. and Snyder, M. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, **148**, 1293-1307.

355. Stafford, P., Wrapp, D. and Johnston, S.A. (2016) General Assessment of Humoral Activity in Healthy Humans. *Molecular & cellular proteomics : MCP*, **15**, 1610-1621.

356. Partinen, M., Kornum, B.R., Plazzi, G., Jennum, P., Julkunen, I. and Vaarala, O. (2014) Narcolepsy as an autoimmune disease: the role of H1N1 infection and vaccination. *Lancet Neurol*, **13**, 600-613.

357. Ching, M.S., Shen, Y., Tan, W.H., Jeste, S.S., Morrow, E.M., Chen, X., Mukaddes, N.M., Yoo, S.Y., Hanson, E., Hundley, R., Austin, C., Becker, R.E., Berry, G.T., Driscoll, K., Engle, E.C., Friedman, S., Gusella, J.F., Hisama, F.M., Irons, M.B., Lafiosca, T., LeClair, E., Miller, D.T., Neessen, M., Picker, J.D., Rappaport, L., Rooney, C.M., Sarco, D.P., Stoler, J.M., Walsh, C.A., Wolff, R.R., Zhang, T., Nasir, R.H., Wu, B.L. and Children's Hospital Boston Genotype Phenotype Study, G. (2010) Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet*, **153B**, 937-947.

358. Kotagal, S. and Broomall, E. (2012) Sleep in children with autism spectrum disorder. *Pediatr Neurol*, **47**, 242-251.

359. Singh, I. and Rose, N. (2009) Biomarkers in psychiatry. *Nature*, **460**, 202-207.

360. Gjerstorff, M.F. and Ditzel, H.J. (2008) An overview of the GAGE cancer/testis antigen family with the inclusion of newly identified members. *Tissue Antigens*, **71**, 187-192.

361. Archelos, J.J., Storch, M.K. and Hartung, H.P. (2000) The role of B cells and autoantibodies in multiple sclerosis. *Annals of neurology*, **47**, 694-706.

362. Hayden, E.C. (2014) Technology: The $1,000 genome. *Nature*, **507**, 294-295.

363. Merl, J., Deeg, C.A., Swadzba, M.E., Ueffing, M. and Hauck, S.M. (2013) Identification of autoantigens in body fluids by combining pull-downs and organic precipitations of intact immune complexes with quantitative label-free mass spectrometry. *Journal of proteome research*, **12**, 5656-5665.

364. Qin, S., Qiu, W., Ehrlich, J.R., Ferdinand, A.S., Richie, J.P., O'Leary M, P., Lee, M.L. and Liu, B.C. (2006) Development of a "reverse capture" autoantibody microarray for studies of antigen-autoantibody profiling. *Proteomics*, **6**, 3199-3209.

365. Mann, M. and Kelleher, N.L. (2008) Precision proteomics: the case for high resolution and high mass accuracy. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 18132-18138.

366. Schweitzer, B. and Kingsmore, S.F. (2002) Measuring proteins on microarrays. *Curr Opin Biotechnol*, **13**, 14-19.

367. Huttenhain, R., Malmstrom, J., Picotti, P. and Aebersold, R. (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr Opin Chem Biol*, **13**, 518-525.

368. Forsström, B., Axnäs, B.B., Stengele, K.P., Buhler, J., Albert, T.J., Richmond, T.A., Hu, F.J., Nilsson, P., Hudson, E.P., Rockberg, J. and Uhlen, M. (2014) Proteome-wide epitope mapping of antibodies using ultra-dense peptide arrays. *Molecular & cellular proteomics : MCP*, **13**, 1585-1597.

369. Liepe, J., Marino, F., Sidney, J., Jeko, A., Bunting, D.E., Sette, A., Kloetzel, P.M., Stumpf, M.P.H., Heck, A.J.R. and Mishto, M. (2016) A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science*, **354**, 354-358.

370. Laumont, C.M., Daouda, T., Laverdure, J.P., Bonneil, E., Caron-Lizotte, O., Hardy, M.P., Granados, D.P., Durette, C., Lemieux, S., Thibault, P. and Perreault, C. (2016) Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nature communications*, **7**.

371. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J. and Mann, M. (2015) Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Molecular & cellular proteomics : MCP*, **14**, 658-673.

372. Bassani-Sternberg, M., Braunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M.E., Werner, A., Hein, R., D, H.B., Peschel, C., Rad, R., Cox, J., Mann, M. and Krackhardt, A.M. (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nature communications*, **7**, 13404.