Biomeditsinskaya Khimiya, 2023 vol. 69, issue 1, pp. 5-18.

REVIEWS

ANTIBODY PROTEOMICS

L.Sh. Kazieva, T.E. Farafonova, V.G. Zgoda*

Institute of Biomedical Chemistry,
10 Pogodinskaya str., Moscow, 119121 Russia; *e-mail: victor.zgoda@gmail.com

Antibodies represent an essential component of humoral immunity; therefore their study is important for molecular biology and medicine. The unique property of antibodies to specifically recognize and bind a certain molecular target (an antigen) determines their widespread application in treatment and diagnostics of diseases, as well as in laboratory and biotechnological practices. High specificity and affinity of antibodies is determined by the presence of primary structure variable regions, which are not encoded in the human genome and are unique for each antibody-producing B cell clone. Hence, there is little or no information about amino acid sequences of the variable regions in the databases. This differs identification of antibody primary structure from most of the proteomic studies because it requires either B cell genome sequencing or *de novo* amino acid sequencing of the antibody. The present review demonstrates some examples of proteomic and proteogenomic approaches and the methodological arsenal that proteomics can offer for studying antibodies, in particular, for identification of primary structure, evaluation of posttranslational modifications and application of bioinformatics tools for their decoding.

Key words: antibodies; mass-spectrometry; *de novo* antibody sequencing; antibody repertoire

DOI: 10.18097/PBMC20236901005

INTRODUCTION

Antibodies are an important component of humoral immunity, and therefore their study is of great importance for molecular biology and medicine. The synthesis of antibodies to various exogenous and endogenous antigens indicates the presence of an immune response, which is an indicator of a particular pathological process in the body. This makes it possible to use antibodies as biomarkers for diagnostics of various infectious, autoimmune, and oncological diseases [1-3]. In addition to diagnostics, antibodies are used as molecular tools for identification, isolation/purification, and labeling of proteins (enzymatic immunoassay, radioimmunoassay, Western blotting, affinity chromatography, cytofluorimetry) [4-7].

The unique ability of antibodies to specifically recognize and bind to a molecular target (an antigen) and subsequently eliminate it, makes it possible to develop drugs for targeted therapy of various diseases.

Understanding molecular mechanisms of the relationship between the specificity of antibodies and their structure will open up new possibilities in the treatment and diagnostics of diseases. The function of antibodies is largely determined by the amino acid sequence of their protein molecule. One of the methods for determining the amino acid sequence of proteins is the method of mass spectrometric analysis, followed by the interpretation of mass spectra by means of databases containing protein sequences. This is possible due to the fact that the genomes of many organisms, including humans, have been already decoded.

The characteristic feature of the antibody structure is the presence of so-called variable regions; their amino acid sequences are unique for each specific B lymphocyte clone [8]. This means that the standard approach for mass spectra identification by existing databases is not applicable. Therefore the sequencing of the variable part of the antibody may be carried out using the genome analysis of activated B lymphocytes, or using amino acid sequencing by mass spectrometric methods.

The aim of this review was to consider the proteomic capacities and the variety of currently existing methodological approaches for studying the structure of antibodies and determination of their complete amino acid sequence, assessing the prospects for the use of antibodies in therapy and diagnostics, as well as identifying the existing problems in the field of proteomic studies of antibodies and possible ways to solve them.

1. STRUCTURE AND FUNCTIONS OF ANTIBODIES

Most mammals produce antibodies of five classes: A, M, G, E and D; the class G immunoglobulins (IgG) are the most abundant in serum. The IgG molecule consists of four polypeptide chains: two heavy chains (HC) and two light chains (LC). Each light chain consists of one variable (VL) and one constant (CL) domains, while the heavy chain consists of one variable (VH) and three constant (CH1-3) domains [9, 10].

In the structure of the antibody, the so-called Fc- (Fragment crystallizable) and Fab- (Fragment antigen-binding) fragments have been recognized (Fig. 1). The Fc fragment consists of CH domains and provides the effector properties of antibodies, while their ability to interact with various antigens is provided by the Fab fragment, namely by its antigen-binding regions (CDR1-3, complementarity determining regions 1-3). Due to the diversity of amino acid sequences in CDRs, antibodies are potentially able to protect the body from external influences, as well as internal disruptions in the genome. The number of possible sequences of the variable domains is huge and, according to some estimates, reaches from 10^8 to 10^{13} variants [11]. The greatest variability is observed in the CDR3 region of the heavy chain (CDRH3). It is assumed that CDRH3 makes the main contribution to the interaction with the antigen [12-14]. For this reason, the attention of many researchers is focused on CDRH3.

In general, understanding the amino acid sequence of the antigen-binding site of a particular antibody is the key step in understanding its function and therapeutic potential.

2. SIGNIFICANCE OF ANTIBODIES IN BIOLOGY AND MEDICINE

Since the discovery of hybridoma technologies, antibodies have become firmly established in medical and scientific practice as therapeutic, diagnostic and laboratory tools. After registration of the first therapeutic monoclonal antibody muromonab (murine monoclonal antibody) in 1986 [15], this group of drugs has become one of the fastest growing areas in the pharmaceutical industry and is widely used in the treatment of various diseases [16]. To date, both monoclonal antibodies and various therapeutic molecules based on them are used in therapy and visualization of pathological processes: antibodies conjugated with active particles — toxins, radioisotopes or cytokines [17], modified antibodies and their fragments — affibodies [18, 19], nanobodies [20], bispecific antibodies [21, 22]. Preparations based on monoclonal antibodies are effective in the treatment of oncological and autoimmune diseases [23-25].

Since antibodies are a component of humoral immunity, their content, as well as changes in the level of their post-translational modifications, make it possible to monitor the reactions of the immune system in the body to external or internal threats. This determines the use of antibodies to exogenous and endogenous antigens as biomarkers of various pathological processes of malignant [26-28], autoimmune [29-32], and infectious nature [33-36], which has become entrenched in clinical practice.

3. ANTIBODY SEQUENCING

The most significant feature of antibodies for proteomics is that the antigen-binding part is not encoded in the genome of the organism, but is determined during recombination of certain DNA regions in each individual B cell and is finally formed as a result of somatic mutations during maturation of activated B lymphocytes [8]. Due to the absence of the desired sequences in genomic

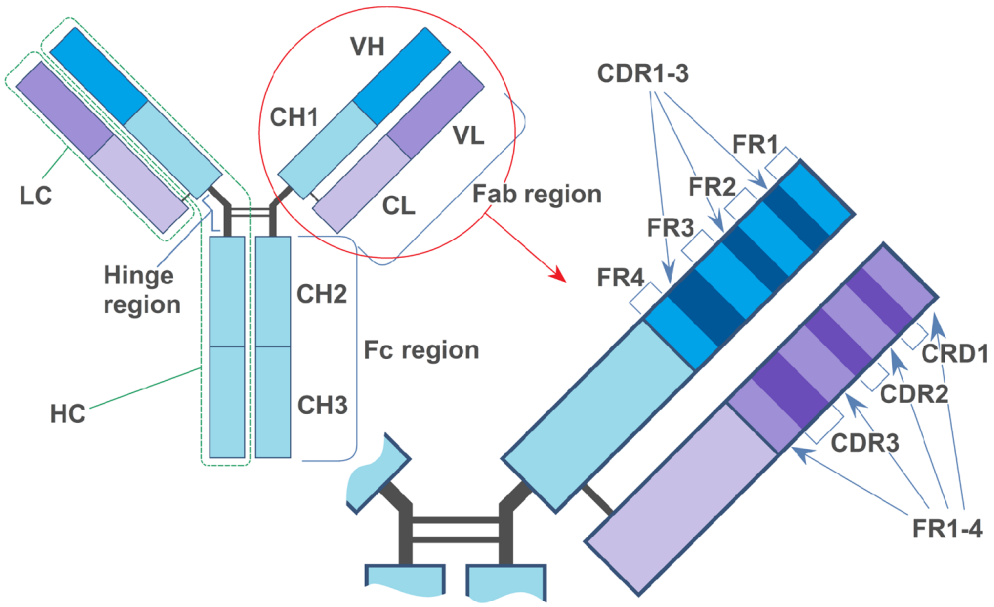

Figure 1. The structure of immunoglobulin G: the antibody domain structure is shown on the left; LC is a light chain, HC is a heavy chain. The light chain consists of the variable (VL) and constant (CL) domains; the heavy chain consists of one variable domain (VH) and three constant domains (CH1-3). On the right, an enlarged diagram of the Fab part shows the detailed structure of the antibody variable domains, having four framework regions (FR1-4) and three hypervariable regions (CDR1-3).

and proteomic databases, the study of the primary structure of antibodies requires information on the corresponding genes that encode them otherwise *de novo* antibody sequencing is needed (Fig. 2).

Thus, determination of variable sequences of antibodies today is possible in two ways:

- genome sequencing of activated B lymphocytes circulating in the blood or B cell receptor (BCR) genes and creating a database of antibody sequences of a particular organism for further proteomic studies.
- *de novo* sequencing of the antibody amino acid sequence by chromato-mass spectrometric methods (LC-MS and LC-MS/MS — chromato-mass spectrometry and tandem chromato-mass spectrometry).

In the arsenal of modern proteomics, three main chromato-mass-spectrometric approaches to the analysis of the primary structure of proteins are currently available: top-down, middle-down, and bottom-up.

The top-down approach implies a mass spectrometric analysis of the whole protein (Fig. 3). The advantages of this approach include the ability to evaluate the molecular weight of an intact protein along with post-translational modifications (PTMs), possible processing of C- and N-terminal amino acids, and also to determine the presence of protein proteoforms [37]. Using different modes of fragmentation of the protein molecule of interest during tandem mass spectrometric analysis (MS/MS), it is possible to obtain the amino acid sequence of the whole antibody. One of the main limitations of the top-down approach is the low performance of the analysis, since the method is not very suitable for working with complex protein mixtures and requires significant amounts of a purified antibody preparation. At the same time, tandem fragmentation spectra, needed to decipher the primary structure of the antibody, almost always require an individual approach.

The bottom-up approach consists in preliminary protein hydrolysis with proteolytic enzymes followed by subsequent chromato-mass-spectrometric (LC-MS/MS) analysis of the resulting mixture of peptides. Peptides have a higher solubility in polar solvents than the whole protein; they are better separated into fractions and ionized and this greatly simplifies their analysis (Fig. 4). For the bottom-up approach, many bioinformatics tools have been developed; they help to identify proteins in a sample by using protein databases, as well as to compile *de novo* amino acid sequences based on MS/MS data [38-45]. One of the main drawbacks of the bottom-up is that some of the peptides obtained during hydrolysis can go beyond the range of detectable mass-to-charge (m/z) values, leading to loss of information about a part of the amino acid

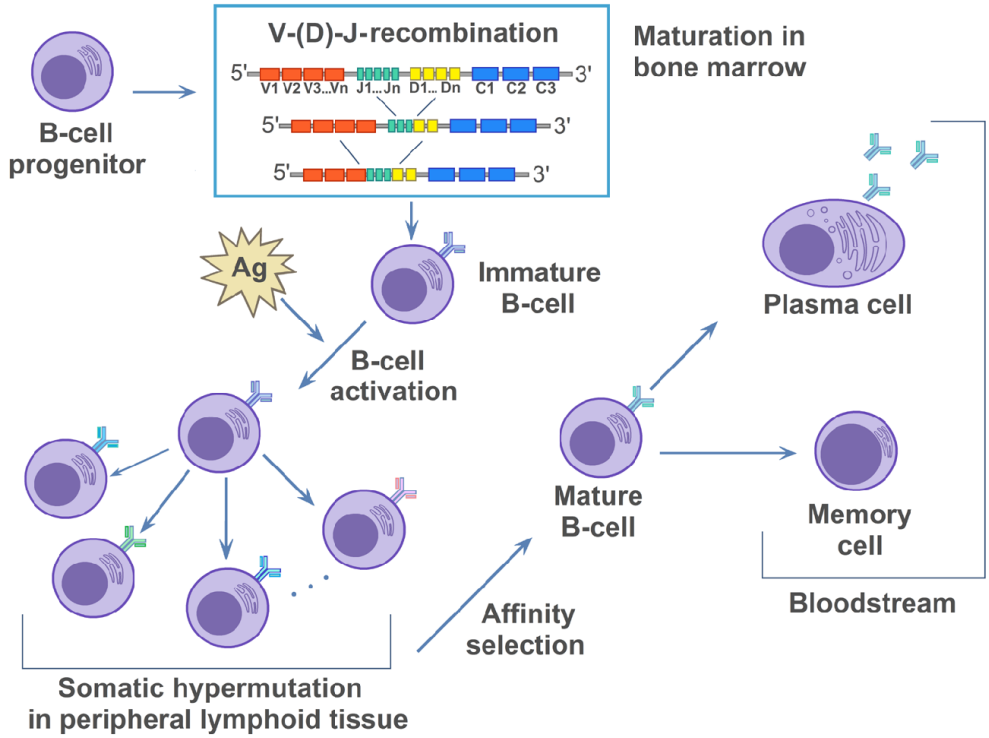

Figure 2. The mechanism providing antibody diversity. In the upper part: B lymphocytes precursor maturation in the bone marrow is accompanied by V-(D)-J-recombination of gene clusters encoding antibody heavy and light chains of antibodies. The resultant immature B cells are activated after contact with an antigen (AG) and migrate to the peripheral organs of the lymphatic system for further maturation. In the lower part of the figure, somatic hypermutation in activated immature B lymphocyte clones makes it possible to achieve a huge variety of possible antibody variants, from which a clone with the highest affinity for AG is selected. This B cell clone is a mature B lymphocyte that further differentiates into active antibody-producing plasma cells or circulating memory B cells.

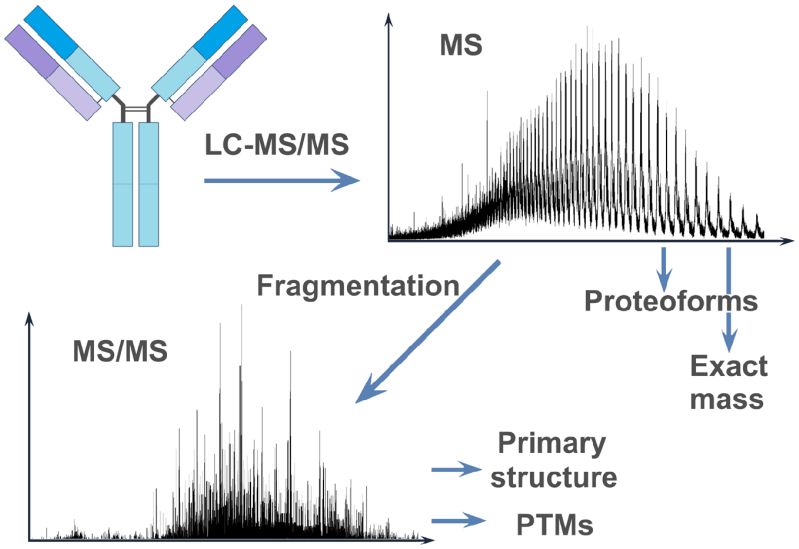

Figure 3. The scheme illustrating the top-down approach. On the top: mass spectrum of the whole antibody, obtained using a high resolution mass spectrometer, is shown. On the bottom: antibody fragmentation to obtain a tandem mass spectrum for determine the antibody primary structure and location of post-translational modifications (PTM) is shown.

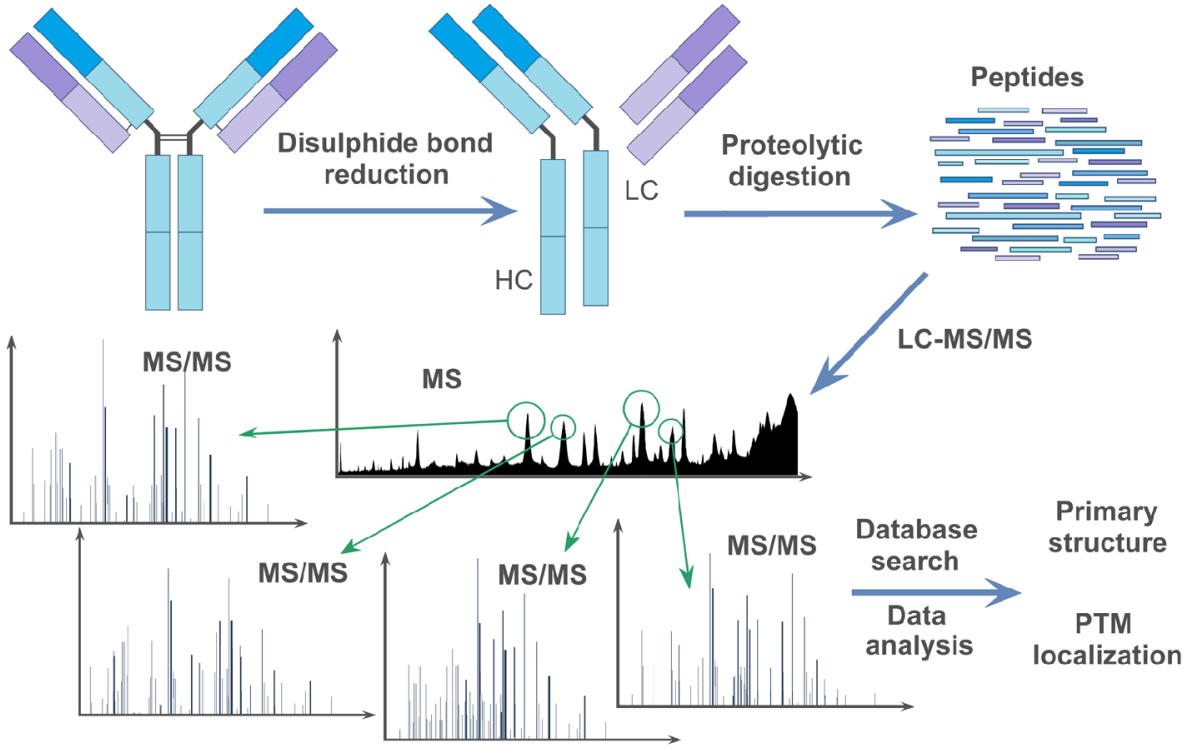

Figure 4. The scheme illustrating the bottom-up approach. On the top: antibody reduction needed for its separation into light (LC) and heavy (HC) chains and subsequent protease hydrolysis. On the bottom: tandem chromatography-mass-spectrometric analysis of the obtained peptides, peptide chromatogram and tandem mass spectra (MS/MS) of their fragments. Using bioinformatics analysis of the obtained data it is possible to determine the primary structure of the studied protein and PTM locations.

sequence and PTM. In the case of antibody sequencing, this is a serious problem, since information about the CDR sequences is critical. Fortunately, this drawback can be overcome by the use of combinations of different proteases (trypsin, chymotrypsin, LysN, LysC, GluC, AspN, thermolysin, elastase): due to their specificity to different cleavage sites, peptides with overlapping sequences can be obtained [46, 47].

The middle-down approach representing a top-down modification is an intermediate version between the top-down and bottom-up (Fig. 5). It includes initial partial cleavage of the protein by S-S bond reduction causing antibody separation into light and heavy chains followed by mass spectrometric analysis. Protein fragments are also obtained by hydrolysis using proteases specific to rare amino acids

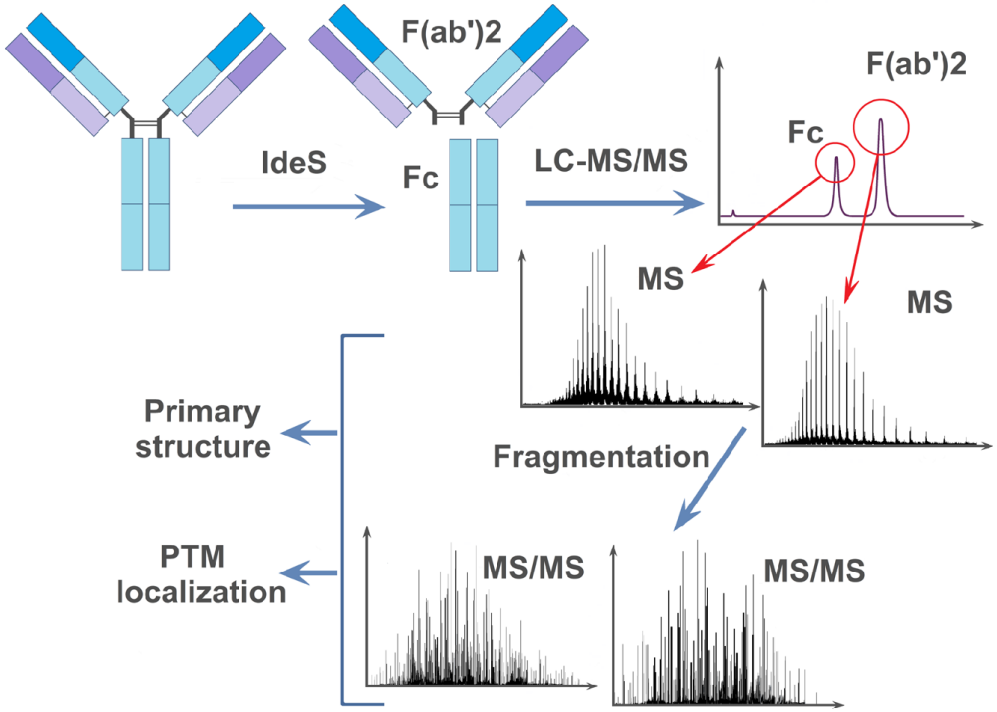

Figure 5. The scheme illustrating the middle-down approach of antibody analysis using the IdeS enzyme. On the top: the IdeS hydrolysis of antibody cleaves the heavy chain at the interface between the hinge region and the CH2 domain. The resulting F(ab′)2 and Fc fragments are analyzed by chromato-mass spectrometry. On the bottom: mass spectra of the F(ab′)2 and Fc parts of the antibody (MS) and the tandem mass spectra of their fragmentation (MS/MS) are shown. Further analysis of the mass spectrometric data makes it possible to determine the antibody primary structure and PTM locations.

(GluC, LysN, LysC, AspN), or to a specific protein site (papain, IdeS), or nonenzymatic hydrolysis in an acidic medium [48]. The middle-down approach is used in studies of various proteins, including antibodies. Due to the fact that sufficiently large antibody fragments are analyzed, the middle-down approach not only increases the degree of their ionization compared to the whole molecule, but also reduces the loss of data on PTM, determines the order of hypervariable regions, and also makes it possible to compare light and heavy chains in mixtures of antibodies [49].

One of the difficulties in protein sequencing by LC-MS/MS consists in the fact that the fragmentation of the parent peptide ions does not discriminate spatial isomers with the same molecular weight, leucine and isoleucine. In the case of *de novo* sequencing of the variable regions of antibodies, the impossibility of reliable determination of all amino acids residues in the sequence creates a problem during the further synthesis of monoclonal antibodies (mAbs) and can lead to a decrease in their affinity. This problem could be solved by combining several methods of ion fragmentation including multi-stage tandem mass spectrometry with high-energy collisional fragmentation (HCD-MS^n) and multi-stage MS/MS (MS^n) with electron transfer fragmentation and analysis of high energy collisional fragmentation (ETD-HCD MS^3) [50].

3.1. Complete Sequencing of Light and Heavy Chains

Complete sequencing of the light and heavy chains of an antibody is necessary when studying modifications of the terminal sequences and PTMs of the antibody, such as glycosylation, methionine oxidation, deamidation of glutamate and aspartate, especially when performing quality control of therapeutic mAbs.

To solve this problem, the top-down mass spectrometric approach is widely used. Mao et al. [51] performed top-down sequencing of native therapeutic mAb by using ECD-MS/MS (electron capture dissociation tandem MS) on a FTICR-MS (Fourier transform ion cyclotron resonance MS) device assembled by the researchers. The authors of the study were able to demonstrate data on 25% coverage of the light and heavy chains during isolation of one charge state precursor; the fragmentation sites were preferentially observed in areas without the S-S linkers. For MS/MS of all charge states without precursor isolation, coverage was 32% and 35% for light and heavy chains, respectively.

Fornelli et al. [52] have demonstrated the use of the middle-down approach for primary IgG sequence analysis using three therapeutic mAbs as an example. For analysis, the antibodies were hydrolyzed with the IdeS protease (*Streptococcus pyogenes* IgG-degrading enzyme); after reduction and alkylation,

~25 kDa fragments were analyzed by chromato-mass spectrometry. Using LC-MS analysis, the exact masses of the antibody fragments after reduction and alkylation were determined, subsequent MS/MS data analysis resulted in determination of the amino acid sequence of the antibody, including hypervariable regions. In addition, the following PTMs were verified: glycosylation sites in the Fc fragment and cleavage of the heavy chain C-terminal lysine. In order to simulate the search for oxidative PTMs using the middle-down approach, one of the antibodies was chemically oxidized under mild conditions prior to IdeS hydrolysis. The researchers were able to identify methionine oxidation sites in the heavy chain. The LC-MS analysis of a mixture of three antibodies hydrolyzed with IdeS was also performed. Despite the chromatographic coelution of several different fragments from different antibodies, their mass spectra made it possible to determine the masses of precursor ions (parent ions), which could be used for further targeted MS/MS analysis and determination of the primary structure of each antibody in the mixture.

In the other study [49], Fornelli et al., using rituximab as the model antibody, showed the possibilities of various fragmentation methods for the top-down and middle-down analysis of antibodies. The monoclonal antibody was analyzed both in native and hydrolyzed form. For hydrolysis, two enzymes, gingipain K and IdeS, were used; the former hydrolyzed the antibody into fragments of 50 kDa while the latter formed the 25 kDa fragments (after reduction and alkylation). Using fragmentation of protein ions by high energy electron transfer dissociation (ETD-HD), ultraviolet photodissociation (UVPD), and electron transfer dissociation/enhanced energy collisional dissociation (EThcD) methods in separate experiments, the researchers obtained a coverage of 40% of the native antibody sequence, 20% during the analysis of fragments weighing 50 kDa (gingipain K), more than 70% for the Fd fragment and up to 90% for LC and Fc/2 (reduced Fc fragment) (IdeS).

In some studies, the top-down approach was used in combination with ultra performance liquid chromatography (1D UPLC-TDMS) for the separation and analysis of Fab fragments of blood serum antibodies [53, 54]. At the initial stage, bottom-up proteomic studies of antibodies usually include creation of a reference genomic database. However, methods based on MS without relying on B lymphocyte genome databases are currently being developed. For example, Cheng et al. [55] demonstrated a sequencing technique based on capillary electrophoresis (CE) of tryptic peptides followed by tandem mass spectrometry using a therapeutic Mab as an example (Fig. 6). The researchers were able to achieve 100% light chain coverage and 96% heavy chain coverage. They also proposed a combination of LC-MS/MS and CE-MS/MS methods. This mixed approach was demonstrated in the sequencing of a new therapeutic anti-CD176 antibody and allowed the authors to be the first to propose an amino acid sequence for this antibody [55].

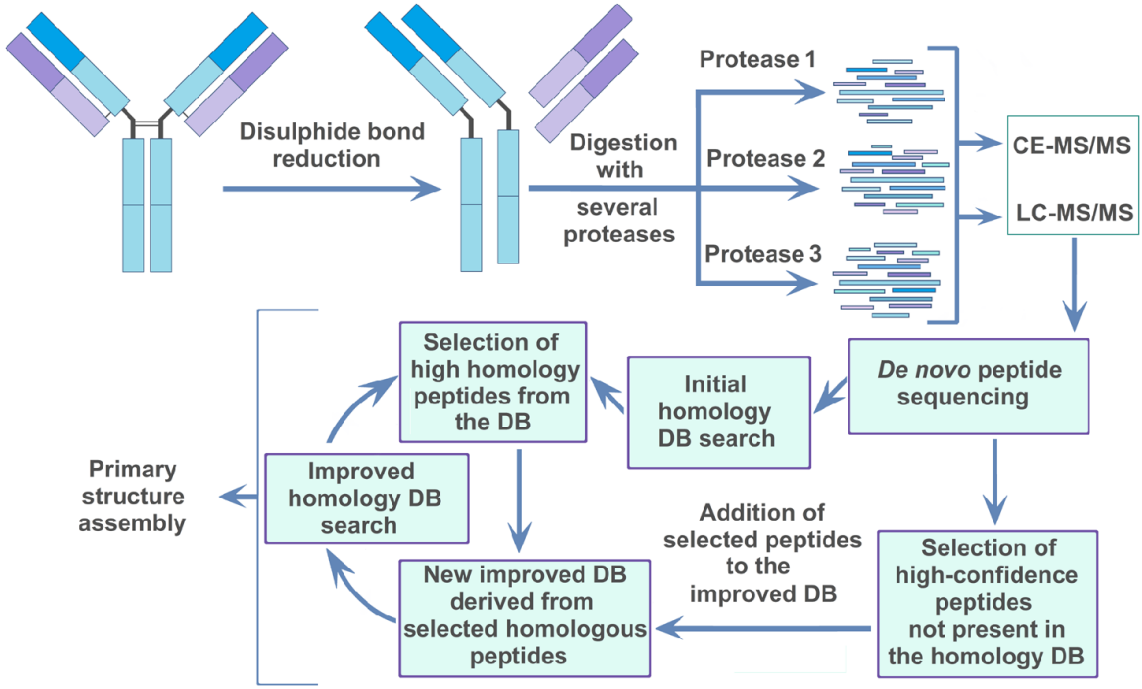

Figure 6. Study design by Cheng et al. [55]. The upper part shows sample preparation and acquisition of experimental data by capillary electrophoresis – tandem mass spectrometry (CE-MS/MS) and tandem chromato-mass spectrometry (LC-MS/MS). The lower part shows the bioinformatic analysis of the obtained data: after constructing the *de novo* sequences and the primary search in a database (DB) of homologous proteins, a new improved database was compiled with the addition of peptides obtained *de novo* with high reliability and not found in the original database. By repeating this database optimization cycle, the amino acid sequence of the test antibody (AB) is compiled.

3.2. Sequencing of Antigen-Binding Sites

Due to the fact that the constant region of antibodies is species-specific, researchers are often interested in sequencing the variable region and CDRs. Sequencing of hypervariable regions of antibodies is also used to study the clonal composition (repertoire) of antibodies in various immune processes.

For example, Wheatley et al. [56] quantified SARS-CoV-2 S protein-specific antibodies and cellular immune responses in a series of blood samples taken from patients during the first 4 months after recovery from COVID-19. The combination of B cell receptor genome sequencing and proteomic methods has identified and quantified circulating antibodies to the S1 subdomain of the S protein in the patients recovered from COVID-19. In this study, the S1-specific IgG was isolated from the plasma of convalescent SARS-CoV-2 patients by means of antigen-coated magnetic particles (AG), and the heavy chains were subjected to LC-MS/MS. In order to determine clonotypes based on the CDRH3 amino acid sequence, peptide spectra were compared with BCR genomic sequences obtained from single memory B cells from the same individuals. The clonotype-specific peptides were then used as barcodes for relative quantitative parallel response monitoring (PRM) and the identified peptides were monitored in plasma samples for 4 months. Quantification of peptides based on chromatographic curves showed a more rapid decline in the neutralizing response during the first 70 days after infection [56].

Lavinder et al. [57] analyzed the antibody repertoire after tetanus revaccination. The researchers isolated antibodies from serum of volunteers after immunization with the vaccine; after antibody hydrolysis into F(ab′)2- and Fc-parts, the AG-binding F(ab′)2-parts were isolated by an affinity-based technique. Using the bottom-up approach, mass spectrometric data of peptides were obtained and the peptides were then identified using individual B cell databases of VH and VL genes of the patients. It was found that the repertoire of antibodies to tetanus toxoid after revaccination consisted of approximately 100 antibody clonotypes, including three clonotypes, which accounted for 40% of the immune response. To check the correctness of sequencing, 13 recombinant antibodies were synthesized; according to the results of affinity evaluation for AG, five synthesized antibodies showed binding to the tetanus toxin epitope, responsible for penetration into the cell. It has also been found that of the diversity of plasma cell clonotypes at the peak of vaccine response, only <5% are detectable at the ninth month post-vaccination. This and other similar studies help to take a deeper look at the mechanisms of the immune response to vaccination, as well as assess the effectiveness of vaccines and the need for their refinement.

Guthals et al. [58] developed a technique for sequencing antibodies to cytomegalovirus (CMV), using bottom-up proteomics and LC-MS of native antibodies. Polyclonal antibodies to CMV were affinity isolated from plasma of an infected patient, subjected to reduction and alkylation. The resulting antibody light and heavy chains were separated by SDS-PAGE and then in-gel digested with various proteases. Light and heavy chain hydrolysates were analyzed by LC-MS/MS. The researchers also measured the exact masses of anti-CMV antibodies after their deglycosylation. The resultant mass spectrometric data was used to construct the amino acid sequence of antibodies *de novo*. Based on the data obtained, recombinant anti-CMV antibodies were synthesized, and their affinity was confirmed by ELISA with an immobilized AG. The researchers found that these anti-CMV antibodies, which accounted for 1-2% of the total IgG, were not detected by parallel sequencing of B cell DNA from the same patient.

The use of a combined proteogenomic approach to investigate antibody repertoires and their dynamic changes may be also illustrated by the study of changes in the IgG repertoire to the influenza virus performed by Lee et al. [59]. They used blood samples repeatedly taken from a patient over several years, both in the stages of an active immune response after vaccination or infection, and during the recovery period. Antibody diversity was determined by the sequences of the CDRH3 heavy chain hypervariable region. To assess the repertoire, antibodies were isolated from the patient's blood by AH (hemagglutinin) affinity chromatography, hydrolyzed with trypsin, and analyzed by chromato-mass spectrometry. An individual database of the patient's peripheral B cell VH genes was used to identify the peptides. After studying the clonal composition of antibodies and the dynamics of its changes during five years, the researchers revealed 210 unique clonotypes. The composition of the polyclonal mixture of anti-hemagglutinin antibodies generally remained basically unchanged, with prevalence of a small number of clonotypes, which persisted steadily during the study period and had significant affinity for highly conserved hemagglutinin regions of the influenza virus.

Bondt et al. [60] studied secretory IgA1 Fab fragments from breast milk obtained from two healthy donors over a period of 16 weeks. Chromato-mass-spectrometric study has shown that the antibody repertoire of each donor does not change, at least for 16 weeks; at the same time, clonal repertoires in two separate donors did not overlap. According to the viewpoint of these authors, the limitation of the middle-down approach is that the real structural and functional uniqueness of each antibody clone is not determined.

3.3. Comparison of Antibody Light and Heavy Chains

Studying polyclonal antibodies, researchers face a number of difficulties that are conceptually different from the study of a monoclonal antibody. Since a polyclonal fraction of B cells is used in genome sequencing, one of these difficulties is the correct matching of antibody light and heavy chains. This problem can be solved by combining genomic and proteomic methods; these include B cell immunoglobulin receptor gene sequencing (BCR) to create a BCR gene database and its use during antibody sequence by means of the bottom-up approach.

In this context, the researchers of the Georgiou group proposed the combined use of mass spectrometry and genomic technologies for antibody sequencing. They applied the developed method, representing a combination of B cell receptor genome sequencing and bottom-up chromato-mass spectrometry for efficient evaluation of serum antibody repertoires after an immune response [61]. The *in vivo* immune response needed to obtain a polyclonal mixture of antibodies was modeled by immunizing rabbits with hemocyanin from the mollusc *Concholepas concholepas*. AG-specific B cells isolated from the bone marrow and blood of immunized rabbits were sequenced, and the resultant gene sequences of the variable region of antibodies were collected in VH- and VL-databases. All IgGs, isolated from blood by affinity chromatography, were hydrolyzed with pepsin, and the AG-specific fraction was isolated from the resulting F(ab′)2 fragments by affinity chromatography with AG. Next, the selected F(ab′)2 fragments were digested with trypsin and the hydrolyzate was analyzed by LC-MS/MS. Peptide identification was performed using the database of VH and VL genes collected by the researchers with the addition of all protein coding sequences of the rabbit genome (the OryCun2 database). During evaluation of the clonal composition, unique peptides, corresponded only to one CDRH3 region, were considered as informative. Comparing genomic and proteomic data, the researchers identified the CDRH3 group, for which frequency of the corresponding peptides were found to be one order of magnitude higher in the AG-specific fraction of F(ab′)2 fragments as compared to the nonspecific fraction. For further pairing of light and heavy chains, seven VH genes containing the most represented CDRH3 sequences were used. For each of the selected VH genes, a library was compiled from combinations with all light chain sequences found in the genomes of AG-specific B cells. The resulting single-chain VH-VL cross-links (scFv, single-chain Fragment variable) were expressed using phage display, and by incubation of phage libraries with immobilized AG (the biopanning method), the scFv with the highest affinity for the antigen was selected. Next, bacteriophages containing crosslinking genes with high affinity for AG were sequenced. Based on the obtained pairs of VH and VL genes, recombinant antibodies were synthesized. To check the correctness of pairs of amino acid sequences of light and heavy chains, the affinity of recombinant antibodies was determined by competitive ELISA for AG.

The use of the middle-down approach is another way to solve this problem. Using this approach it is possible to analyze large protein fragments under non-reduced conditions. In this case the light chain and the heavy chain fragment containing the VH domain are not uncoupled in the sample and therefore this approach is applicable for the study non-reduced F(ab′)2 fragments of antibodies [48]. Results of several studies [49, 62, 63] show, that the use of the middle-down approach, especially in combination with other chromato-mass-spectrometric approaches and with several methods of ion fragmentation, makes it possible to reliably obtain information about light-heavy chain pairs in the structure of antibodies.

3.4. Features of Data Processing for Antibody Sequencing

For efficient *de novo* sequencing of mono- and polyclonal antibodies, not only experimental approaches to sequencing and mass spectrometry, but also effective bioinformatic tools for data processing that take into consideration specific structure of antibodies, are of great importance. During optimization of the analysis of proteomic and proteogenomic data in the study of antibodies researchers meet a number of problems including:

- reduction in the number of false positive identifications of peptides during the analysis of mass spectra of antibody peptides;
- optimization and automation of the processing of mass spectrometric data of antibody peptides during sequencing;
- development of methods for *de novo* sequencing of antibodies that do not require creation of a database of B cell receptor genes, but are based on bioinformatic reconstruction of the V-D-J recombination process;
- the variable region of the antibody, particularly CDRs, is shorter than the constant region, therefore, during antibody hydrolysis, peptides of the variable region will be shorter than the constant region peptides. This means that methods are needed to maximize the acquisition of information about all the peptides present in the sample.

The study by Boutz et al. [64] represents a good example of solving one of these problems. The authors sequenced the VH and VL genes of B cells of immunized rabbits (AG, *Concholepas concholepas* mollusc hemocyanin) and humans (AG, tetanus/diphtheria toxoid). Antibodies isolated from serum by affinity chromatography with protein A (protein G for human antibodies) were digested with pepsin, and the resulting F(ab′)2 fragments were

subjected to trypsinolysis followed by peptide analysis by LC-MS/MS. The main task of the work was to develop a method for reducing the number of false positive identifications of CDRH3 peptides. The presence of framework regions in the structure of Fab fragments, which have a significantly lower variability compared to CDRs, leads to the fact that the proteolytic cleavage of a mixture of polyclonal Fab fragments yields a pool of peptides that give a high score of coincidence with sequences of several V-genes of B lymphocytes at once. This phenomenon creates difficulties in the analysis of peptide mass spectra, since the standard error modeling method using randomly generated sequences (decoy sequences) does not take into account this feature of antibodies and leads to a high level of false positive identifications. Therefore, it is necessary to use methods with strict selection for the analysis of peptide spectra. To determine the percentage of incorrect identifications and to find characteristics by which true and false matches can be distinguished, the authors of [64] used differential labeling of cysteine residues. This made it possible to compile a list of identifications with the highest correspondence to the labeling data and to identify false positive identifications with a high score. Samples of rabbit F(ab′)2 fragments were divided into two parts and alkylated with two different agents, iodoacetamide and iodoethanol. As a result, both samples contained peptides identical in amino acid sequence, but with a mass difference of 13 Da for modified cysteine residues, which served to determine true identifications. After analyzing the lists of false and true identifications with a high score, the authors were able to achieve an increase in the reliability of peptide identification due to more stringent parameters of the average mass deviation. They showed that at a mean deviation threshold of 1 ppm, an 87% improvement in search accuracy was observed without loss of true identifications [64].

During identification of the results of an LC-MS/MS experiment on protein databases, researchers often face the fact that a part of the good quality mass spectra remains unidentified. This leads to the fact that some proteins or peptides present in the sample may remain unnoticed. To increase the efficiency of mass spectrum analysis, Lundström et al. [65] proposed a new approach, known as SpotLight Proteomics. The SpotLight approach is conceptually similar to the standard bottom-up analysis of antibodies, but has a number of features: antibody peptides are analyzed by LC-MS/MS using two different fragmentation methods (HCD and ETD); based on the interpretation of the obtained MS data using *de novo* algorithms, a database of antibody sequences of the sample is created; subsequently, both the original serum sample and its purified fraction with a high content of IgG are analyzed; for peptide identification by mass spectra, the researchers combined the reference human proteome (Uniprot) and *de novo* decoded sequences into a single SpotLight database. Based on the obtained database, peptides were identified in all samples and quantified using the DeMix-Q algorithm [66]. This approach has been used to compare antibody profiles in patients with early stages of Alzheimer's disease and dementia with Lewy bodies. In the original serum proteome, 156 proteins were quantified, in the proteome of the IgG-enriched sample 81 proteins were identified, and 23 proteins were not detected in the original serum sample. At the level of peptides in intact serum, 18% of 2112 unique sequences were associated with IgG; in the sample enriched with antibodies, the proportion of peptides homologous to IgG was already 31% of 2077 sequences. In the original samples, four proteins and 47 peptides differed significantly in patients with Alzheimer's disease and dementia with Lewy bodies; in the IgG-enriched samples, a difference in quantity was observed for 49 peptides, 23 of which belonged to proteins that differed significantly in abundance in the compared groups.

Sen et al. [45] have demonstrated their Supernovo algorithm, which automates the sequencing of monoclonal antibodies (Fig. 7). This algorithm uses MS/MS data of antibody peptides obtained by the bottom-up method using several proteases. The first step is to search for peptides in the IMGT (International Information System ImMunoGeneTics) database of germline immunoglobulin sequences and subsequent recombination of V-, J-, and C-exons *in silico* to create a template of light and heavy chain sequences. Further, the resulting template is refined by comparing it with the results of *de novo* sequencing of peptides by mass spectra and its correction using the Byonic software package [67]. As a result of several repetitions of this process, the final antibody sequence is assembled.

Automation of the antibody sequencing process is of great importance for pharmacy and biochemistry, since it will reduce the time and financial costs at this stage in the development of new therapeutic, diagnostic, and laboratory MAbs.

Recently, Schulte et al. proposed the Stitch algorithm to optimize data processing in antibody sequencing [68]. This algorithm uses MS/MS spectra passed *de novo* peptide sequences for their mapping by means of customizable amino acid sequence templates (Fig. 8). Using the IMGT database, peptides are grouped according to their light and heavy chains using local alignment. IMGT is also used in this algorithm to create custom templates that align short sequences obtained *de novo* from LC-MS/MS experiments. To reconstruct the sequence of the CDRH3 hypervariable region, the algorithm matches the overlapping regions of the V- and J-parts, aligning them to user template sequences, thus reconstructing the hypervariable regions not covered by the database.

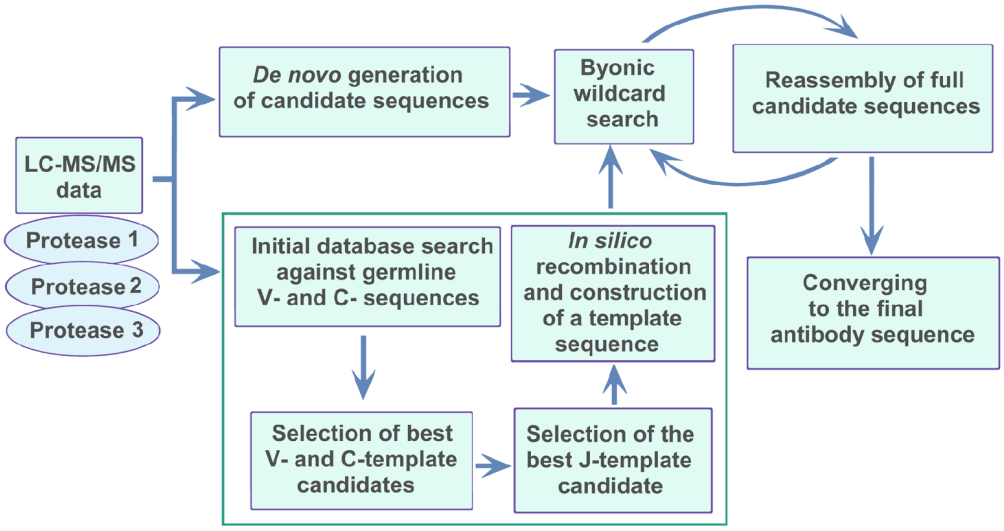

Figure 7. Design of the study by Sen et al. [45]. From left to right, according to the obtained experimental data, *de novo* peptide sequence generation and a search in the ImMunoGeneTics (IMGT) germline immunoglobulin gene database were carried out simultaneously to select template sequences of V- and C-segments with the highest homology to the studied antibody. After J-segment selection to the selected V- and C-, *in silico* recombination was performed to compile a template of the V-J-C segment of the antibody sequence. Based on the obtained templates and *de novo* sequences, candidate peptide sequences were compiled using the Byonic algorithm [67]. After several rounds of adjustment of the candidate peptides, the complete antibody sequence was compiled.

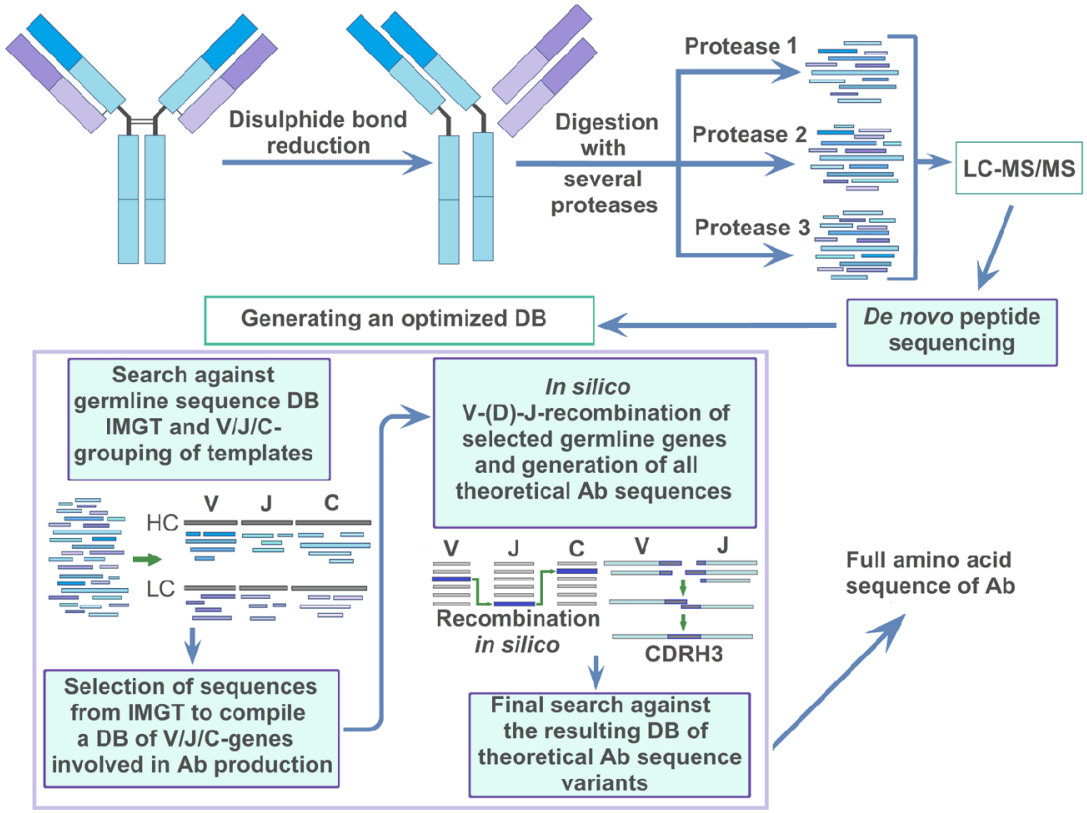

Figure 8. Design of the study by Schulte et al. [68]. The upper part shows sample preparation and tandem chromatography-mass-spectrometric analysis of antibody peptides. The lower part shows the bioinformatic analysis of the obtained mass spectra. An optimized database (DB) was created to identify the peptides and compile the antibody (AB) sequence:for *de novo* peptides obtained, a search was performed in the ImMunoGeneTics (IMGT) germline immunoglobulin gene database; the AB peptides were sorted by their correspondence to V-, J-, and C - sites of heavy (HC) and light (LC) chains. The sequences from the IMGT database most homologous to the studied peptides were selected for *in silico* modeling of V-(D)-J recombination, reconstruction of the CDRH3 heavy chain hypervariable region, and creation of a database of theoretical AB sequences. According to the resulting database, a repeated search was carried out to compile the AB amino acid sequence.

Using this methodology, the researchers demonstrated the applicability of Stitch by sequencing two monoclonal antibodies, a single Fab fragment in a mixture of plasma polyclonal antibodies, sequencing light chains in the urine of patients with multiple myeloma, and assessing the antibody repertoire in people hospitalized with COVID-19.

CONCLUSIONS

Antibodies are crucially important for modern molecular biology, medicine, and pharmacy. Their unique properties, which allow specific binding of a particular molecular target, stimulated the development of many technologies and techniques used in therapy, diagnostics, drug production, and biomedical laboratories. LC-MS/MS has proven to be a highly efficient method for the complete amino acid sequencing of monoclonal antibodies and determining the presence of post-translational modifications. The study of the antibody diversity in polyclonal mixtures is a more difficult task due to the small proportion of sequences that differ between clonotypes in the antibody molecule. At the same time, the use of only blood B cell BCR gene sequencing is not sufficiently informative, since peripheral B cells do not reflect the repertoire of antibodies in the blood. This makes mass spectrometric analysis an indispensable method in the study of polyclonal antibodies.

Technological development of mass spectrometric equipment expands the possibilities of using MS for *de novo* sequencing of antibodies, particularly, due to the possibility of combining several methods of ion fragmentation in one instrument. Although the efficient fragmentation of a whole antibody or its large parts still represents a difficult task, it is likely that further technological progress will sufficiently improve mass spectrometers to introduce top-down and middle-down approaches into routine proteomic practice together with the bottom-up approach.

FUNDING

The work was performed within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021-2030) (No. 122030100168-2).

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any research involving humans or using animals as subjects.

CONFLICT OF INTERESTS

The authors declare no conflict of interests.

REFERENCES

1. *Yman V., Tuju J., White M.T., Kamuyu G., Mwai K., Kibinge N., Asghar M., Sundling C., Sondén K., Murungi L, Kiboi D., Kimathi R., Chege T., Chepsat E., Kiyuka P., Nyamako L., Osier F.H.A., Färnert A.* (2022) Distinct kinetics of antibodies to 111 *Plasmodium falciparum* proteins identifies markers of recent malaria exposure. Nat. Commun., **13**(1), 331-331. DOI: 10.1038/s41467-021-27863-8
2. *Monroy-Iglesias M.J., Crescioli S., Beckmann K., Le N., Karagiannis S.N., van Hemelrijck M., Santaolalla A.* (2022) Antibodies as biomarkers for cancer risk: A systematic review. Clin. Exper. Immunol., **209**(1), 46-63. DOI: 10.1093/cei/uxac030
3. *Satoh M., Ceribelli A., Hasegawa T., Tanaka S.* (2022) Clinical significance of antinucleolar antibodies: Biomarkers for autoimmune diseases, malignancies, and others. Clin. Rev. Allergy Immunol., **63**(2), 210-239. DOI: 10.1007/s12016-022-08931-3
4. *Alsaed O.S., Alamlih L.I., Al-Radideh O., Chandra P., Alemadi S., Al-Allaf A.-W.* (2021) Clinical utility of ANA-ELISA vs ANA-immunofluorescence in connective tissue diseases. Sci. Rep., **11**(1), 8229-8229. DOI: 10.1038/s41598-021-87366-w
5. *Musharova O., Medvedeva S., Klimuk E., Guzman N.M., Titova D., Zgoda V., Shiriaeva A., Semenova E., Severinov K., Savitskaya E.* (2021) Prespacers formed during primed adaptation associate with the Cas1-Cas2 adaptation complex and the Cas3 interference nuclease-helicase. Proc. Nat. Acad. Sci. USA, **118**(22), e2021291118. DOI: 10.1073/pnas.2021291118
6. *Neubert H., Shuford C.M., Olah T.V., Garofolo F., Schultz G.A., Jones B.R., Amaravadi L., Laterza O.F., Xu K., Ackermann B.L.* (2020) Protein biomarker quantification by immunoaffinity liquid chromatography-tandem mass spectrometry: Current state and future vision. Clinical Chemistry, **66**(2), 282-301. DOI: 10.1093/clinchem/hvz022
7. *Brummelman J., Haftmann C., Núñez N.G., Alvisi G., Mazza E.M.C., Becher B., Lugli E.* (2019) Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry. Nature Protocols, **14**(7), 1946-1969. DOI: 10.1038/s41596-019-0166-2
8. *David M.P.C., Asprer J.J.T., Ibana J.S.A., Concepcion G.P., Padlan E.A.* (2007) A study of the structural correlates of affinity maturation: antibody affinity as a function of chemical interactions, structural plasticity and stability. Mol. Immunol., **44**(6), 1342-1351. DOI: 10.1016/J.MOLIMM.2006.05.006
9. *Stanfield R.L., Wilson I.A.* (2014) Antibody structure. Microbiology Spectrum, **2**(2), DOI: 10.1128/microbiolspec.AID-0012-2013
10. *Snapkov I., Chernigovskaya M., Sinitcyn P., Lê Quý K., Nyman T.A., Greiff V.* (2022) Progress and challenges in mass spectrometry-based analysis of antibody repertoires. Trends Biotechnol., **40**(4), 463-481. DOI: 10.1016/J.TIBTECH.2021.08.006
11. *Greiff V., Miho E., Menzel U., Reddy S.T.* (2015) Bioinformatic and statistical analysis of adaptive immune repertoires. Trends Immunol., **36**(11), 738-749. DOI: 10.1016/j.it.2015.09.006
12. *Xu J.L., Davis M.M.* (2000) Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity, **13**(1), 37-45. DOI: 10.1016/S1074-7613(00)00006-6

13. *Khass M., Vale A.M., Burrows P.D., Schroeder H.W.* (2018) The sequences encoded by immunoglobulin diversity (DH) gene segments play key roles in controlling B-cell development, antigen-binding site diversity, and antibody production. Immunol. Rev., **284**(1), 106-119. DOI: 10.1111/imr.12669
14. *Robinson S.A., Raybould M.I.J., Schneider C., Wong W.K., Marks C., Deane C.M.* (2021) Epitope profiling using computational structural modelling demonstrated on coronavirus-binding antibodies. PLoS Comput. Biol., **17**(12), e1009675. DOI: 10.1371/journal.pcbi.1009675
15. *Todd P.A., Brogden R.N.* (1989) Muromonab CD3. Drugs, **37**(6), 871-899. DOI: 10.2165/00003495-198937060-00004
16. *Lu R.-M., Hwang Y.-C., Liu I.-J., Lee C.-C., Tsai H.-Z., Li H.-J., Wu H.-C.* (2020) Development of therapeutic antibodies for the treatment of diseases. J. Biomed. Sci., **27**(1), 1-1. DOI: 10.1186/s12929-019-0592-z
17. *Awwad S., Angkawinitwong U.* (2018) Overview of antibody drug delivery. Pharmaceutics, **10**(3), 83-83. DOI: 10.3390/pharmaceutics10030083
18. *Frejd F.Y., Kim K.-T.* (2017) Affibody molecules as engineered protein drugs. Exper. Mol. Med., **49**(3), e306-e306. DOI: 10.1038/emm.2017.35
19. *Brasino M., Roy S., Erbse A.H., He L., Mao C., Park W., Cha J.N., Goodwin A.P.* (2018) Anti-EGFR affibodies with site-specific photo-cross-linker incorporation show both directed target-specific photoconjugation and increased retention in tumors. J. Am. Chem. Soc., **140**(37), 11820-11828. DOI: 10.1021/jacs.8b07601
20. *Altunay B., Morgenroth A., Beheshti M., Vogg A., Wong N.C.L., Ting H.H., Biersack H.-J., Stickeler E., Mottaghy F.M.* (2021) HER2-directed antibodies, affibodies and nanobodies as drug-delivery vehicles in breast cancer with a specific focus on radioimmunotherapy and radioimmunoimaging. Eur. J. Nucl. Med. Mol. Imaging, **48**(5), 1371-1389. DOI: 10.1007/s00259-020-05094-1
21. *Brinkmann U., Kontermann R.E.* (2021) Bispecific antibodies. Science, **372**(6545), 916-917. DOI: 10.1126/science.abg1209
22. *Lee S.C., Ma J.S.Y., Kim M.S., Laborda E., Choi S.-H., Hampton E.N., Yun H., Nunez V., Muldong M.T., Wu C.N., Ma W., Kulidjian A.A., Kane C.J., Klyushnichenko V., Woods A.K., Joseph S.B., Petrassi M., Wisler J., Li J. et al* (2021) A PSMA-targeted bispecific antibody for prostate cancer driven by a small-molecule targeting ligand. Science Advances, **7**(33), eabi8193. DOI: 10.1126/sciadv.abi8193
23. *Oldham R.K., Dillman R.O.* (2008) Monoclonal antibodies in cancer therapy: 25 years of progress. J. Clin. Oncol., **26**(11), 1774-1777. DOI: 10.1200/JCO.2007.15.7438
24. *Hafeez U., Gan H.K., Scott A.M.* (2018) Monoclonal antibodies as immunomodulatory therapy against cancer and autoimmune diseases. Curr. Opin. Pharmacol., **41**, 114-121. DOI: 10.1016/j.coph.2018.05.010
25. *Melero I., Hervas-Stubbs S., Glennie M., Pardoll D.M., Chen L.* (2007) Immunostimulatory monoclonal antibodies for cancer therapy. Nature Reviews. Cancer, **7**(2), 95-106. DOI: 10.1038/nrc2051
26. *Fujiwara S., Wada H., Kawada J., Kawabata R., Takahashi T., Fujita J., Hirao T., Shibata K., Makari Y., Iijima S. et al.* (2013) NY-ESO-1 antibody as a novel tumour marker of gastric cancer. Br. J. Cancer, **108**(5), 1119-1125. DOI: 10.1038/bjc.2013.51
27. *Lee Z.J.O., Eslick G.D., Edirimanne S.* (2020) Investigating antithyroglobulin antibody as a prognostic marker for differentiated thyroid cancer: A meta-analysis and systematic review. Thyroid, **30**(11), 1601-1612. DOI: 10.1089/thy.2019.0368
28. *Bogdanos D.P., Gkoutzourelas A., Papadopoulos V., Liaskos C., Patrikiou E., Tsigalou C., Saratziotis A., Hajiioannou J., Scheper T., Meyer W., Sakkas L.I., Papandreou C.* (2021) Anti-Ro52 antibody is highly prevalent and a marker of better prognosis in patients with ovarian cancer. Clin. Chim. Acta, **521**, 199-205. DOI: 10.1016/j.cca.2021.07.006
29. *Scherer H.U., Huizinga T.W.J., Krönke G., Schett G., Toes R.E.M.* (2018) The B cell response to citrullinated antigens in the development of rheumatoid arthritis. Nature Reviews. Rheumatology, **14**(3), 157-169. DOI: 10.1038/nrrheum.2018.10
30. *Darrah E., Giles J.T., Ols M.L., Bull H.G., Andrade F., Rosen A.* (2013) Erosive rheumatoid arthritis is associated with antibodies that activate PAD4 by increasing calcium sensitivity. Sci. Transl. Med., **5**(186), 186ra65. DOI: 10.1126/scitranslmed.3005370
31. *Zhou X., Motta F., Selmi C., Ridgway W.M., Gershwin M.E., Zhang W.* (2021) Antibody glycosylation in autoimmune diseases. Autoimmun. Rev., **20**(5), 102804-102804. DOI: 10.1016/j.autrev.2021.102804
32. *Fechtner S., Berens H., Bemis E., Johnson R.L., Guthridge C.J., Carlson N.E., Demoruelle M.K., Harley J.B., Edison J.D., Norris J.A., Robinson W.H., Deane K.D., James J.A., Holers V.M.* (2022) Antibody responses to Epstein-Barr virus in the preclinical period of rheumatoid arthritis suggest the presence of increased viral reactivation cycles. Arthritis Rheumatology, **74**(4), 597-603. DOI: 10.1002/art.41994
33. *Lu L.L., Chung A.W., Rosebrock T.R., Ghebremichael M., Yu W.H., Grace P.S., Schoen M.K., Tafesse F., Martin C., Leung V. et al.* (2016) A functional role for antibodies in tuberculosis. Cell, **167**(2), 433-443.e14. DOI: 10.1016/j.cell.2016.08.072
34. *Alter G., Ottenhoff T.H.M., Joosten S.A.* (2018) Antibody glycosylation in inflammation, disease and vaccination. Semin. Immunol., **39**, 102-110. DOI: 10.1016/j.smim.2018.05.003
35. *Offersen R., Yu W.-H., Scully E.P., Julg B., Euler Z., Sadanand S., Garcia-Dominguez D., Zheng L., Rasmussen T.A., Jennewein M.F. et al.* (2020) HIV antibody Fc N-linked glycosylation is associated with viral rebound. Cell Reports, **33**(11), 108502-108502. DOI: 10.1016/j.celrep.2020.108502
36. *Bournazos S., Vo H.T.M., Duong V., Auerswald H., Ly S., Sakuntabhai A., Dussart P., Cantaert T., Ravetch J.V.* (2021) Antibody fucosylation predicts disease severity in secondary dengue infection. Science, **372**(6546), 1102-1105. DOI: 10.1126/science.abc7303
37. *Toby T.K., Fornelli L., Srzentić K., de Hart C.J., Levitsky J., Friedewald J., Kelleher N.L.* (2019) A comprehensive pipeline for translational top-down proteomics from a single blood draw. Nature Protocols, **14**(1), 119-152. DOI: 10.1038/s41596-018-0085-7
38. *Hirosawa M., Hoshida M., Ishikawa M., Toya T.* (1993) MASCOT: Multiple alignment system for protein sequences based on three-way dynamic programming. Bioinformatics, **9**(2), 161-167. DOI: 10.1093/bioinformatics/9.2.161
39. *Tyanova S., Temu T., Cox J.* (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nature Protocols, **11**(12), 2301-2319. DOI: 10.1038/nprot.2016.136
40. *Barsnes H., Vaudel M.* (2018) SearchGUI: A highly adaptable common interface for proteomics search and *de novo* engines. J. Proteome Res., **17**(7), 2552-2555. DOI: 10.1021/acs.jproteome.8b00175

41. *Vaudel M., Burkhart J.M., Zahedi R.P., Oveland E., Berven F.S., Sickmann A., Martens L., Barsnes H.* (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. Nature Biotechnology, **33**(1), 22-24. DOI: 10.1038/nbt.3109
42. *Zhang J., Xin L., Shan B., Chen W., Xie M., Yuen D., Zhang W., Zhang Z., Lajoie G.A., Ma B.* (2012) PEAKS DB: *De novo* sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics, **11**(4), M111.010587-M111.010587. DOI: 10.1074/mcp.M111.010587
43. *Frank A., Pevzner P.* (2005) PepNovo: *De novo* peptide sequencing via probabilistic network modeling. Anal. Chem., **77**(4), 964-973. DOI: 10.1021/ac048788h
44. *Jeong K., Kim S., Pevzner P.A.* (2013) UniNovo: A universal tool for *de novo* peptide sequencing. Bioinformatics, **29**(16), 1953-1962. DOI: 10.1093/bioinformatics/btt338
45. *Sen K.I., Tang W.H., Nayak S., Kil Y.J., Bern M., Ozoglu B., Ueberheide B., Davis D., Becker C.* (2017) Automated antibody *de novo* sequencing and its utility in biopharmaceutical discovery. J. Am. Soc. Mass Spectrom., **28**(5), 803-810. DOI: 10.1007/s13361-016-1580-0
46. *Zhang Y., Fonslow B.R., Shan B., Baek M.-C., Yates J.R.* (2013) Protein analysis by shotgun/bottom-up proteomics. Chem. Rev., **113**(4), 2343-2394. DOI: 10.1021/cr3003533
47. *Peng W., Pronker M.F., Snijder J.* (2021) Mass spectrometry-based *de novo* sequencing of monoclonal antibodies using multiple proteases and a dual fragmentation scheme. J. Proteome Res., **20**, 3559-3566. DOI: 10.1021/acs.jproteome.1c00169
48. *Cristobal A., Marino F., Post H., van den Toorn H.W.P., Mohammed S., Heck A.J.R.* (2017) Toward an optimized workflow for middle-down proteomics. Anal. Chem., **89**(6), 3318-3325. DOI: 10.1021/acs.analchem.6b03756
49. *Fornelli L., Srzentić K., Huguet R., Mullen C., Sharma S., Zabrouskov V., Fellers R.T., Durbin K.R., Compton P.D., Kelleher N.L.* (2018) Accurate sequence analysis of a monoclonal antibody by top-down and middle-down orbitrap mass spectrometry applying multiple ion activation techniques. Anal. Chem., **90**(14), 8421-8429. DOI: 10.1021/acs.analchem.8b00984
50. *Xiao Y., Vecchi M.M., Wen D.* (2016) Distinguishing between leucine and isoleucine by integrated LC-MS analysis using an orbitrap fusion mass spectrometer. Anal. Chem., **88**(21), 10757-10766. DOI: 10.1021/acs.analchem.6b03409
51. *Mao Y., Valeja S.G., Rouse J.C., Hendrickson C.L., Marshall A.G.* (2013) Top-down structural analysis of an intact monoclonal antibody by electron capture dissociation-fourier transform ion cyclotron resonance-mass spectrometry. Anal. Chem., **85**(9), 4239-4246. DOI: 10.1021/ac303525n
52. *Fornelli L., Ayoub D., Aizikov K., Beck A., Tsybin Y.O.* (2014) Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap fourier transform mass spectrometry. Anal. Chem., **86**(6), 3005-3012. DOI: 10.1021/ac4036857
53. *Wang Z., Liu X., Muther J., James J.A., Smith K., Wu S.* (2019) Top-down mass spectrometry analysis of human serum autoantibody antigen-binding fragments. Sci. Rep., **9**(1), 2345. DOI: 10.1038/s41598-018-38380-y
54. *Rosati S., Yang Y., Barendregt A., Heck A.J.R.* (2014) Detailed mass analysis of structural heterogeneity in monoclonal antibodies using native mass spectrometry. Nature Protocols, **9**(4), 967-976. DOI: 10.1038/nprot.2014.057
55. *Cheng J., Wang L., Rive C.M., Holt R.A., Morin G.B., Chen D.D.Y.* (2020) Complementary methods for *de novo* monoclonal antibody sequencing to achieve complete sequence coverage. J. Proteome Res., **19**(7), 2700-2707. DOI: 10.1021/acs.jproteome.0c00223
56. *Wheatley A.K., Juno J.A., Wang J.J., Selva K.J., Reynaldi A., Tan H.-X., Lee W.S., Wragg K.M., Kelly H.G., Esterbauer R., Davis S.K., Kent H.E., Mordant F.L., Schlub T.E., Gordon D.L., Khoury D.S., Subbarao K., Cromer D., Gordon T.P., Chung A.W., Davenport M.P., Kent S.J.* (2021) Evolution of immune responses to SARS-CoV-2 in mild-moderate COVID-19. Nat. Commun., **12**(1), 1162-1162. DOI: 10.1038/s41467-021-21444-5
57. *Lavinder J.J., Wine Y., Giesecke C., Ippolito G.C., Horton A.P., Lungu O.I., Hoi K.H., de Kosky B.J., Murrin E.M., Wirth M.M., Ellington A.D., Dörner T., Marcotte E.M., Boutz D.R., Georgiou G.* (2014) Identification and characterization of the constituent human serum antibodies elicited by vaccination. Proc. Nat. Acad. Sci. USA, **111**(6), 2259-2264. DOI: 10.1073/pnas.1317793111
58. *Guthals A., Gan Y., Murray L., Chen Y., Stinson J., Nakamura G., Lill J.R., Sandoval W., Bandeira N.* (2017) *De novo* MS/MS sequencing of native human antibodies. J. Proteome Res., **16**(1), 45-54. DOI: 10.1021/acs.jproteome.6b00608
59. *Lee J., Paparoditis P., Horton A.P., Frühwirth A., McDaniel J.R., Jung J., Boutz D.R., Hussein D.A., Tanno Y., Pappas L., Ippolito G.C., Corti D., Lanzavecchia A., Georgiou G.* (2019) Persistent antibody clonotypes dominate the serum response to influenza over multiple years and repeated vaccinations. Cell Host Microbe, **25**(3), 367-376.e5. DOI: 10.1016/j.chom.2019.01.010
60. *Bondt A., Dingess K.A., Hoek M., van Rijswijck D.M.H., Heck A.J.R.* (2021) A direct MS-based approach to profile human milk secretory immunoglobulin A (IgA1) reveals donor-specific clonal repertoires with high longitudinal stability. Front. Immunol., **12**, 789748. DOI: 10.3389/fimmu.2021.789748
61. *Wine Y., Boutz D.R., Lavinder J.J., Miklos A.E., Hughes R.A., Hoi K.H., Jung S.T., Horton A.P., Murrin E.M., Ellington A.D., Marcotte E.M., Georgiou G.* (2013) Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. Proc. Nat. Acad. Sci., **110**(8), 2993-2998. DOI: 10.1073/pnas.1213737110
62. *Shaw J.B., Liu W., Vasil'ev Y.V., Bracken C.C., Malhan N., Guthals A., Beckman J.S., Voinov V.G.* (2020) Direct determination of antibody chain pairing by top-down and middle-down mass spectrometry using electron capture dissociation and ultraviolet photodissociation. Anal. Chem., **92**(1), 766-773. DOI: 10.1021/acs.analchem.9b03129
63. *Srzentić K., Nagornov K.O., Fornelli L., Lobas A.A., Ayoub D., Kozhinov A.N., Gasilova N., Menin L., Beck A., Gorshkov M.V., Aizikov K., Tsybin Y.O.* (2018) Multiplexed middle-down mass spectrometry as a method for revealing light and heavy chain connectivity in a monoclonal antibody. Anal. Chem., **90**(21), 12527-12535. DOI: 10.1021/acs.analchem.8b02398
64. *Boutz D.R., Horton A.P., Wine Y., Lavinder J.J., Georgiou G., Marcotte E.M.* (2014) Proteomic identification of monoclonal antibodies from serum. Anal. Chem., **86**(10), 4758-4766. DOI: 10.1021/ac4037679
65. *Lundström S.L., Zhang B., Rutishauser D., Aarsland D., Zubarev R.A.* (2017) SpotLight proteomics: Uncovering the hidden blood proteome improves diagnostic power of proteomics. Sci. Rep., **7**(1), 41929-41929. DOI: 10.1038/srep41929

66. *Zhang B., Pirmoradian M., Chernobrovkin A., Zubarev R.A.* (2014) DeMix workflow for efficient identification of cofragmented peptides in high resolution data-dependent tandem mass spectrometry. Mol. Cell. Proteomics, **13**(11), 3211-3223. DOI: 10.1074/mcp.O114.038877
67. *Bern M., Kil Y.J., Becker C.* (2012) Byonic: Advanced peptide and protein identification software. Curr. Protoc. Bioinformatics, **40**(1), 13.20.1-13.20.14. DOI: 10.1002/0471250953.bi1320s40
68. *Schulte D., Peng W., Snijder J.* (2022) Template-based assembly of proteomic short reads for *de novo* antibody sequencing and repertoire profiling. Anal. Chem., **94**(29), 10391-10399. DOI: 10.1021/acs.analchem.2c01300

Received: 08. 11. 2022.
Revised: 26. 12. 2022.
Accepted: 28. 12. 2022.

ПРОТЕОМИКА АНТИТЕЛ

Л.Ш. Казиева, Т.Е. Фарафонова, В.Г. Згода*

Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича, 119121, Москва, Погодинская ул., 10; *эл. почта: victor.zgoda@gmail.com

Антитела являются важным компонентом гуморального иммунитета, в связи с чем их изучение имеет колоссальное значение для молекулярной биологии и медицины. Уникальная способность антител специфично распознавать и связываться с молекулярной мишенью — антигеном — обусловливает их широкое применение в терапии и диагностике заболеваний, а также в лабораторной и биотехнологической практике. Высокая специфичность и аффинность антител обеспечивается наличием в их структуре вариабельных участков, которые не закодированы в геноме человека, а уникальны для каждого конкретного клона В-лимфоцитов — продуцентов антител. Как следствие, информация о последовательностях вариабельных участков антител в базах данных почти отсутствует. Это отличает изучение первичной структуры антител от других протеомных исследований, поскольку подразумевает расшифровку генома В-лимфоцитов либо *de novo* секвенирование аминокислотной последовательности антител. В данном обзоре на примере ряда протеомных и протео-геномных исследований рассмотрены существующие подходы, которые протеомика может предложить для изучения антител, в частности, для расшифровки их первичной структуры, оценки посттрансляционных модификаций и оптимизации биоинформатических инструментов с учётом специфики анализа антител.

Полный текст статьи на русском языке доступен на сайте журнала (http://pbmc.ibmc.msk.ru).

Ключевые слова: антитела; масс-спектрометрия; секвенирование антител *de novo*; репертуар антител

Финансирование. Работа выполнена в рамках Программы фундаментальных научных исследований в Российской Федерации на долгосрочный период (2021-2030 годы) (№ 122030100168-2).

Поступила в редакцию: 08.11.2022; после доработки: 26.12.2022; принята к печати: 28.12.2022.