CAPILLARY ELECTROPHORESIS FOR MODERN BIOPHARMACEUTICALS: MASTERING LARGE MOLECULES AND SMALL DETAILS

INAUGURALDISSERTATION

ZUR

ERLANGUNG DER WÜRDE EINES DOKTORS DER PHILOSOPHIE

VORGELEGT DER

PHILOSOPHISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER UNIVERSITÄT BASEL

VON

ANDREI HUTANU

BASEL, 2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Erstbetreuerin:	PD. Dr. Maria A. Schwarz
Zweitbetreuer:	Prof. Dr. Peter C. Hauser
externer Experte :	Prof. Dr. Hermann Wätzig

Basel, den 17 Oktober 2023

Prof. Dr. Marcel Mayor

Dekan

ABSTRACT

The purpose of this thesis is to investigate the utility of capillary electrophoresis (CE) as an analytical technique for assessing the critical quality attributes of novel therapeutic agents for their use as innovative drugs, such as peptides, plasmids, mRNA/lipid nano particle formulations, and recombinant adeno-associated virus (rAAV). The primary objective is to enhance CE's applicability to these modalities by developing novel approaches for these challenging therapeutics. The dissertation incorporates four projects that were covered in a total of five articles published in peer-reviewed analytical chemistry periodicals. The first chapter investigates the ability of affinity CE to quantify the degree and location of methionine oxidation in peptides and monoclonal antibodies with the aim of providing a complementary approach to mass spectrometry. The following chapter provides a direct comparison of the present analytical methodologies for the analysis of rAAV including transmission electron microscopy, analytical ultra centrifugation, capillary gel electrophoresis (CGE), capillary zone electrophoresis (CZE) and ion-pair reverse phase liquid chromatography (IP-RP-LC). The third chapter describes a method for quantifying and identifying nucleic acid therapeutics using fluorescent peptide nucleic acid (PNA) hybridization probes in CE. The final part of this work reviews various analytical methods for the analysis of high-mass oligonucleotides that are incorporated in different delivery vehicles, highlighting particularly methodological benefits and drawbacks. This includes the PNA-based CE method from the previous chapter, digital droplet PCR, CGE from the second chapter with the additional and complementary use of nuclease digestion, IP-RP-LC, and CZE which were all applied for the comprehensive characterization of oligonucleotides and their quality attributes as biologics.

“Happiness lies in the joy of achievement, in the thrill of creative effort.”

Franklin Delano Roosevelt

Table of Contents

1. Introduction

Biopharmaceuticals are a broad category of pharmaceutical products, mostly parenterals, that include any biological drug created by living organisms using recombinant DNA technology. Prominent examples are recombinant protein therapeutics, vaccines, and cell- and gene-therapies. [1] While recombinant therapeutics did not reach patients until the 1980s, biological products such as porcine insulin and poliovirus vaccines were already well established. [2] A majority of these early products were derived from animal tissues or serum. The inherent variability in these manufacturing platforms resulted in low product purity and efficacy when compared to modern standards. [3] Growing understanding of production techniques has played a significant role in the rapid advancement of biopharmaceuticals and making them available for patients as new therapeutic approaches, addressing unmet medical needs. [4] The process operations used today have been influenced by technologies developed in the first half of the twentieth century that are based on fundamental chemical principles such as chromatography and electrophoresis. Incremental advancements in these technologies have resulted in present biopharma techniques that have transcended consistency and reproducibility, resulting in enhanced product efficacy, quality, and safety.

Today, more than 400 distinct active biopharmaceuticals are available, with 197 regulatory approvals between 2018 and mid-2022. [5] The biological nature of these products allows for the development of novel mechanisms of action that would be impossible to achieve with small molecules, which is exploited for the treatment of a wide range of indications such as cancer, autoimmune diseases, genetic disorders or infectious diseases. Although a broad variety of modalities (e.g. hormones, growth factors, enzymes and viral vectors) is available for patients, the vast majority (> 80%) of commercial biopharmaceuticals are monoclonal antibodies (mAbs). [5] The ability of antibodies to effectively bind to specific targets with a strong affinity and specificity, and their potential to activate the immune system based on the desired functionality, has contributed to their widespread use as therapeutics. This is particularly advantageous compared to small molecules, which face difficulties in achieving these tasks. [6] The high demand for therapeutical mAbs has driven much of the biopharma industry's advances in manufacturing and process improvement, with mAb research results being applied to other products. [4] A typical example is the prosperity of cell and gene therapies (C&GT). After more than forty years of research, multiple such therapies were approved by the regulatory authorities over the last decade. [5] Just as mAbs revolutionized the treatment of various diseases by offering novel

mechanisms of action, C&GTs have the potential to offer personalized treatment options and long-term or even permanent cures for previously difficult-to-treat diseases. Similarly to mAbs, the current demand for C&GTs is also anticipated to spur additional developments in the sector. [7]

1.1 Analytical Method Development for Biopharmaceuticals

A biopharma product's overall life cycle begins with discovery, when a variety of initial candidate molecules are generated and tested for activity against a specific biological target. The best candidates are optimized to improve efficacy and safety. Products selected for further human testing are assigned to process development, where the manufacturing processes required to supply material in sufficient quantities for clinical trials and, in case successful clinical trials, commercial needs are developed. [4]

The development and application of reliable analytical techniques is critical throughout the entire lifecycle (Figure 1). During the early stages of discovery and optimization, analytical methods must be developed to characterize candidate molecules and measure their activity against the biological target. The results obtained by these techniques are then used to guide candidate selection, pre-formulation screening and later to release drug substance or drug product that is required for non-clinical toxicological studies. [8] This first stage is, therefore, characterized by the focus to understand the drug substance. Hereafter, analytical development works in tandem with process development to improve the process's quality and integrity while maximizing efficiency and cost-effectiveness. Monitoring the process is critical to ensure that the final product is consistent and reproducible, but also in ensuring high product quality and safety. [9]

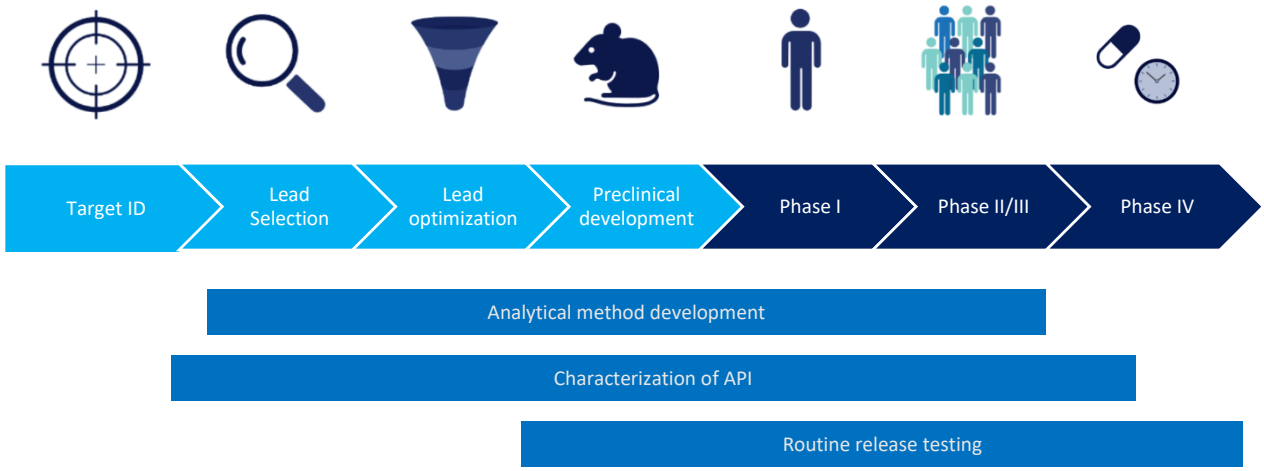


Figure 1: Typical pharmaceutical development lifecycle and the role of analytical methods.

Finally, upon completion of the development phase, validation or verification of the chosen analytical techniques is performed in order to obtain evidence that the methods are fit for intended use. Validated methods are mandatory for quality control, release, and stability testing of drug substance and drug product for clinical studies and later on, if market authorization has been granted, during commercial manufacturing. [8]

The objective of analytical development is to establish a framework for precise and reliable surveillance and verification of novel molecular entities throughout their complete life cycle, specifically by developing robust analytical methodologies. This development process starts by defining the purpose of the method and the conditions in which it should be employed. For these quality attributes the product specific requirements for method parameters such as accuracy, precision, sensitivity etc. are defined. [10] Based on the assumptions made in the initial step a technology platform that can achieve all criteria is selected. For common modalities such as mAbs a panel of platform methods including widely accepted generic criteria is typically available, whereas a complete method development might be necessary for emerging therapeutic entities such as C&GT products. [8] After selecting the analytical technology platform for a new test method and establishing the basic method parameters, systematic method evaluation and development can be performed to assess the potential impact of the various method parameters on the method's ability to consistently produce robust, accurate, and precise results. This step may require further fine tuning of the method attributes in order to meet the initial purpose. As validation of the method after completion of the development incorporates significant investment, it is closely tied to the lifecycle of the drug candidate. For this reason, the extent of the validation effort is determined by the method type and clinical phase of the program. [9]

1.2 Quality Attributes of Biopharmaceuticals

Each biopharmaceutical has an individual set of attributes that are critical to safeguard the desired product quality. These critical quality attributes (CQAs) are defined early on in the product lifecycle and may be associated with any property (e.g. physical, chemical, biological, immunological etc. etc.) of the final drug. A manufacturer is defining a set of attributes and claims and supports these with appropriate test method results before submitting these to the regulatory authorities. [11,12] Further, CQAs are constantly reviewed as the development of the product progresses if they still cover all aspects and meet the need to control the product quality comprehensively.

CQAs can be classified into various categories depending on the nature of the drug product, the manufacturing process, and other factors. According to Geigert [13] there are eight major classes: Appearance, Identity, Purity, Potency, Quantity, Adventitious Agent Safety, General and Device Functionality. Typical categories (excluding device functionality) and some related examples are shown in Figure 2. Variation in the chemical properties which are linked to a CQA may affect the drug's stability, resulting in changes in therapeutic potency. For example, if the drug's efficacy decreases too much due to instability over time, the patient may not receive the anticipated therapeutic benefit. Other modifications may also have an impact on the drug's safety. Impurities or contaminants may have toxic side effects on the patient's health if they are introduced.

This work focusses on analytical test methods that address CQAs related to purity and identity. Identity testing is used to ensure that the active ingredient and drug product are what they are claimed to be and that they meet the specifications for the intended product. Modern biopharma manufacturing relies on global production networks which include significant logistical efforts. Identity testing ensures the use of the anticipated material for production, i.e. prevents mix-up of products and is thus crucial for product quality and patient safety. [14] The identity of a biopharmaceutical is defined via unique properties that are specific to the active ingredient or product. The sequence of the protein or nucleic acid (NA), post-translational modifications (PTM), glycosylation patterns, and other physicochemical properties are typical examples of these properties. [13] Purity refers to the proportion of the active ingredient in the product compared to related substances and/or impurities, e.g. caused by degradation or compounds that were incorporated during manufacturing. A biopharmaceutical product's purity is critical because impurities or other compounds in the product can have a negative impact on its safety, efficacy, and tolerability. A prominent example is represented by mAb aggregation which are of heightened interest to the health authorities due to their ability to cause strong cytokine and co-stimulatory signals. [15,16] The higher the purity of the product, the lower the risk of side effects or other adverse events. [13] There is a broad variety of impurities that can be part of a drug product. Generally, they can be categorized in two categories: process-related and product-related. [17]

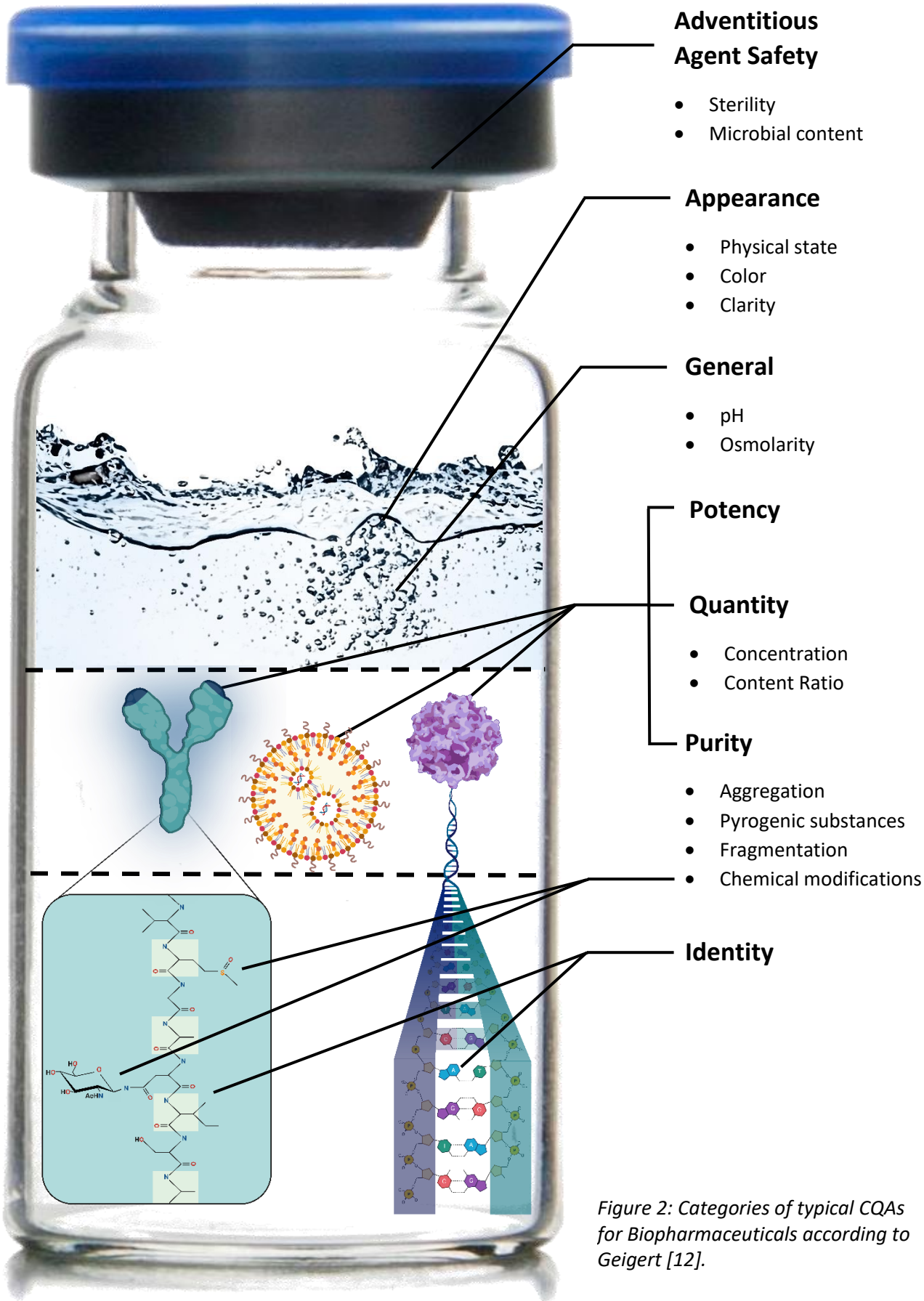


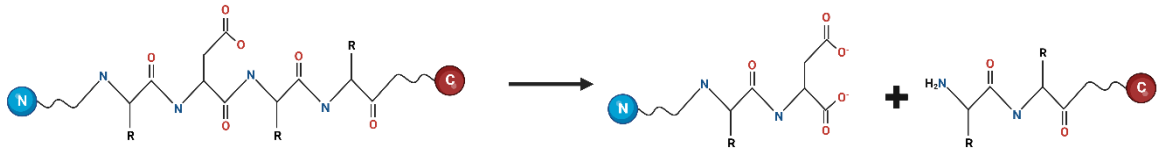
Figure 2: Categories of typical CQAs for Biopharmaceuticals according to Geigert [12].

1.2.1 Quality attributes of protein pharmaceuticals

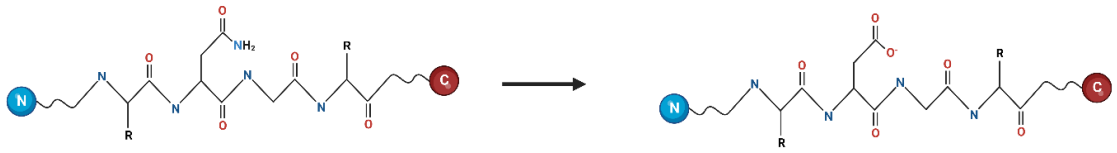
As the name suggests, process-related impurities are derived from the cell culture and purification process and may include chemicals used in the media for active pharmaceutical ingredient (API) production such as antibiotics, processing agents from purification or impurities introduced during filling and packaging. [18] Host cell-related contaminants are typical process related impurities and represent proteins as well as DNA of the host organism that are unintentionally released e.g. into the cell culture medium. Although the medium or cell lysate may contain thousands of proteins or nucleic acids, subsequent purification steps will significantly reduce the number of host cell contaminants during the process. Nevertheless, even impurities present in the ppm range can have significant impacts on the product. An example are lipases which can degrade polysorbate a prominent surfactant. [19] At such low levels further reduction of impurities is difficult and standard analytical techniques often reach their limits, so special workarounds need to be found. These low level components may additionally cause direct side effects to the patients as pyrogenics or they may interact with the API or the formulation and therefore limit stability and effectiveness. [18]

Due to the complex manufacturing process in living cells, biopharmaceuticals are often heterogenous and contain various slight variations of the API. In addition, all these variants can undergo various biochemical and physical degradation processes. Proteins are vulnerable to a number of chemical degradation pathways, such as fragmentation, deamidation, isomerization, oxidation, deglycosylation, glycation, and cross-linking (Figure 3; Figure 4). [20] These PTMs have the potential to alter the biological activity of the therapeutic agent, cause immunogenicity, and promote further degradation. Gao et al. reported that oxidation of a single methionine amino acid in the Fc region of mAbs (Met252 or Met428) can affect binding to the neonatal Fc receptor, thereby decreasing the antibody's half-life. [21] Hydrophobicity, charge heterogeneity, and conformational changes caused by chemical modifications to amino acid residues can all result in the therapeutic's physical instability. Deamidation, for example, can cause a change in a product's charge distribution, which can result in aggregation of a drug product at low pH. [22] Oxidation of Met and Trp can cause local conformational changes in the tertiary structure, promoting aggregation. [23] Although the peptide bond is stable under most conditions, several factors including pH, solvent, temperature, intensity of light exposure, and excipients may accelerate fragmentation. For mAbs several prominent cleavage sites were identified which are prone to hydrolysis and should be taken into account when designing the formulation. [24] Because of their low dissociation energy compared to C-H and C-C bonds, disulfides participate in a variety of modification processes such as fragmentation, oligomerization, aggregation, disulfide

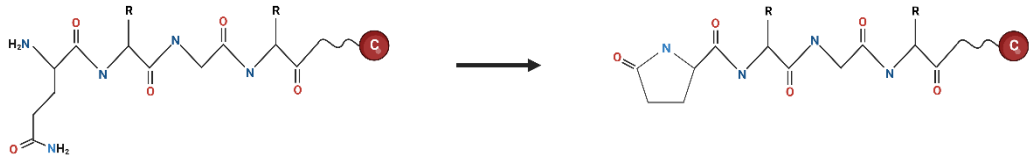
A) Fragmentation => size variants



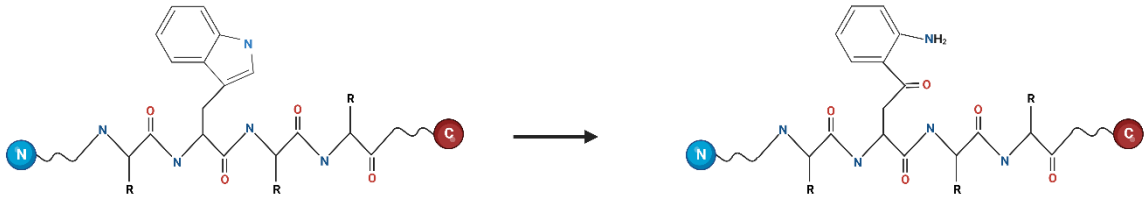
B) Deamidation => charge variants



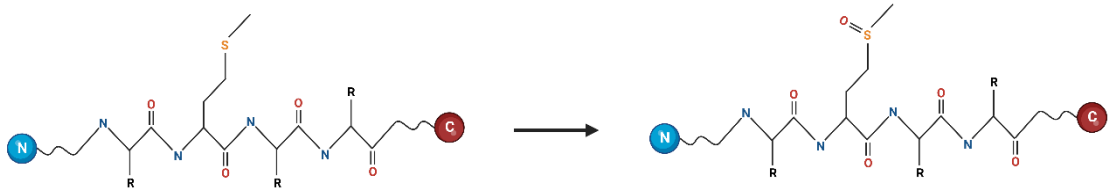
C) Pyroglutamic acid formation => charge variants



D) Trp oxidation => hydrophobic variants



E) Met oxidation => hydrophobic variants



F) Maillard reaction/Glycation => Glycosylation pattern

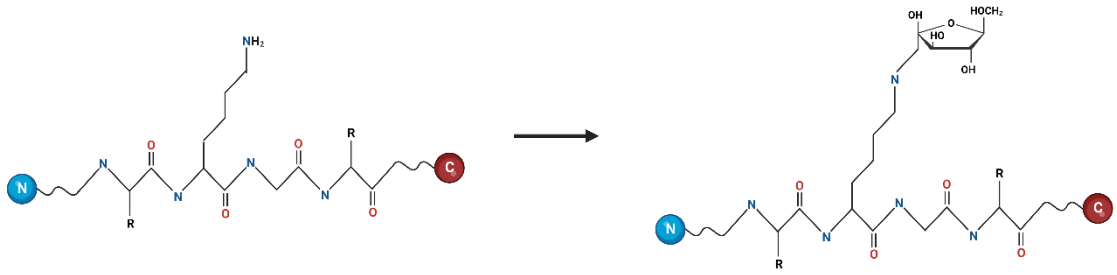
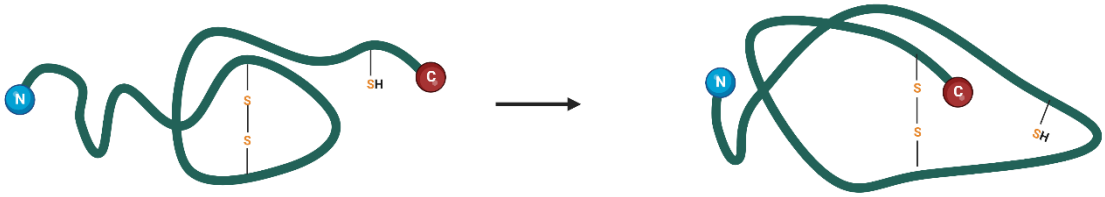
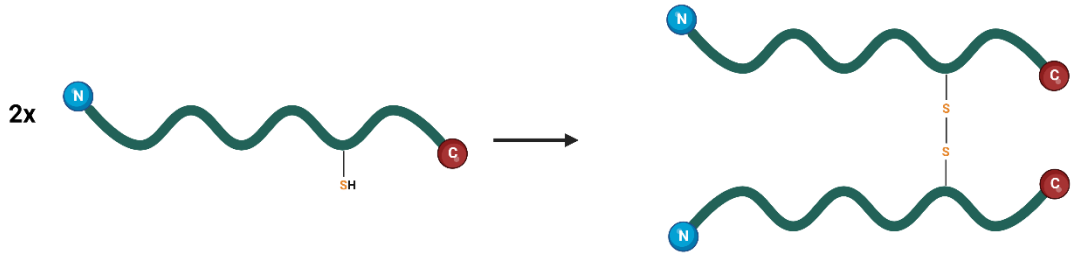


Figure 3: Post-translational modifications of proteins that affect a single amino acid.

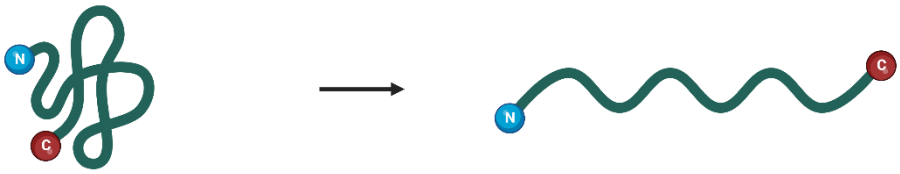
A) Disulfide Exchange => hydrophobic or charge variants



B) Oligomerization => size variants



C) Denaturation => hydrophobic or charge variants



D) Aggregation => size variants; particles

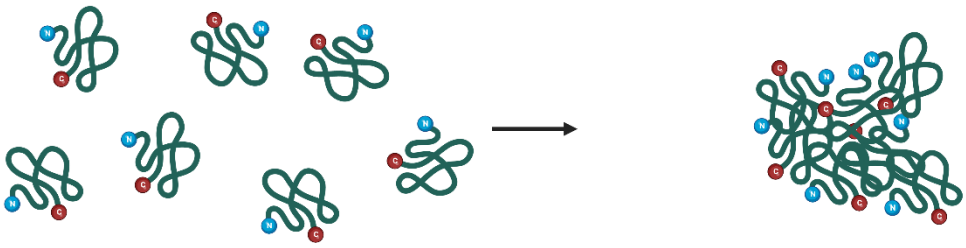


Figure 4: Changes in the higher order structure of protein biopharmaceuticals.

cleavage and disulfide bond exchange. [25] For instance, in trastuzumab, one of the early anti-cancer antibody-based drugs, the loss of a single disulfide bond can lead to a significant loss in anti-tumor efficacy. [26] In general, chemical changes can cause physical instabilities, which in turn can cause chemical changes in amino acid residues, demonstrating that the two phenomena are inextricably linked.

1.2.2 Quality attributes of nucleic acid pharmaceuticals

The picture is similar for therapeutics containing nucleic acids (NAs): Chemical modifications have the potential to greatly impact the efficacy of the drug. However, the molecular environment of these biopharmaceuticals is very diverse, ranging from oligonucleotides that are manufactured by chemical synthesis over plasmids to mRNA, which are produced biotechnologically. Additionally, NAs might be administered on their own or, more often, with a delivery system such as Lipid Nano Particles (LNPs) or viral vectors. [27]

All these variables lead to very different process-related and product-related contaminations. For example mRNA is not produced in cells [28] and thus there is no problem of host-cell impurities, while DNA plasmids are manufactured in bacteria [29] which can pose a great risk for immunogenicity. However, the most challenging class of NA therapeutics are those that incorporate a viral vector which consists out of proteins, potentially lipids and NA. In this case, all problems that might occur with protein therapeutics (*vide supra*) need to be considered in addition to specific NA impurities. The field of NA modifications has been extensively researched in the context of ageing and cancer research. [30-32] By this, a large variety of chemical modifications was discovered which probably play an important role for stability of NA therapeutics, too (Figure 5). [33] However, the application of this knowledge on actual drug products is currently at a very early stage and it is not yet completely clear which effects play superior roles for quality control. Similarly to proteins, NAs are vulnerable to a number of chemical degradation pathways that result in modifications, such as fragmentation, deamination, oxidation, depurination, and changes in the secondary/tertiary structure. Oxidation can happen in all five natural bases but is most often observed at guanosine sites. [34] The formation of 8-Oxo-2'-deoxyguanosine is linked to increased oxidative stress *in vivo* and ultimately with various diseases. Modified guanosine can not only form hydrogen bridges with cytosine but also with adenine, causing mutations and is additionally prone to further oxidation. [35]

DNA Methylation is an *in vivo* mechanism to regulate protein expression. Adenine and Cytosine can be methylated, with cytosine methylation being more widespread. Methylation is not a “modification” *per se* but represents the characteristics of the production process. Typically, methylated DNA acts to repress transcription and can thus be critical for efficacy of DNA therapeutics. [36] On the other hand, hypomethylation can pose risks of immunogenicity. The ability of the vertebrate immune system to identify the presence of bacterial DNA is based on

A) Oxidation: Guanosine to 8-Oxo-2'-deoxyguanosine

B) Methylation: Cytosine to 5-Methylcytosine

C) Deamination: 5-Methylcytosine to Thymine

D) Depurination: apurinic sites

E) Fragmentation

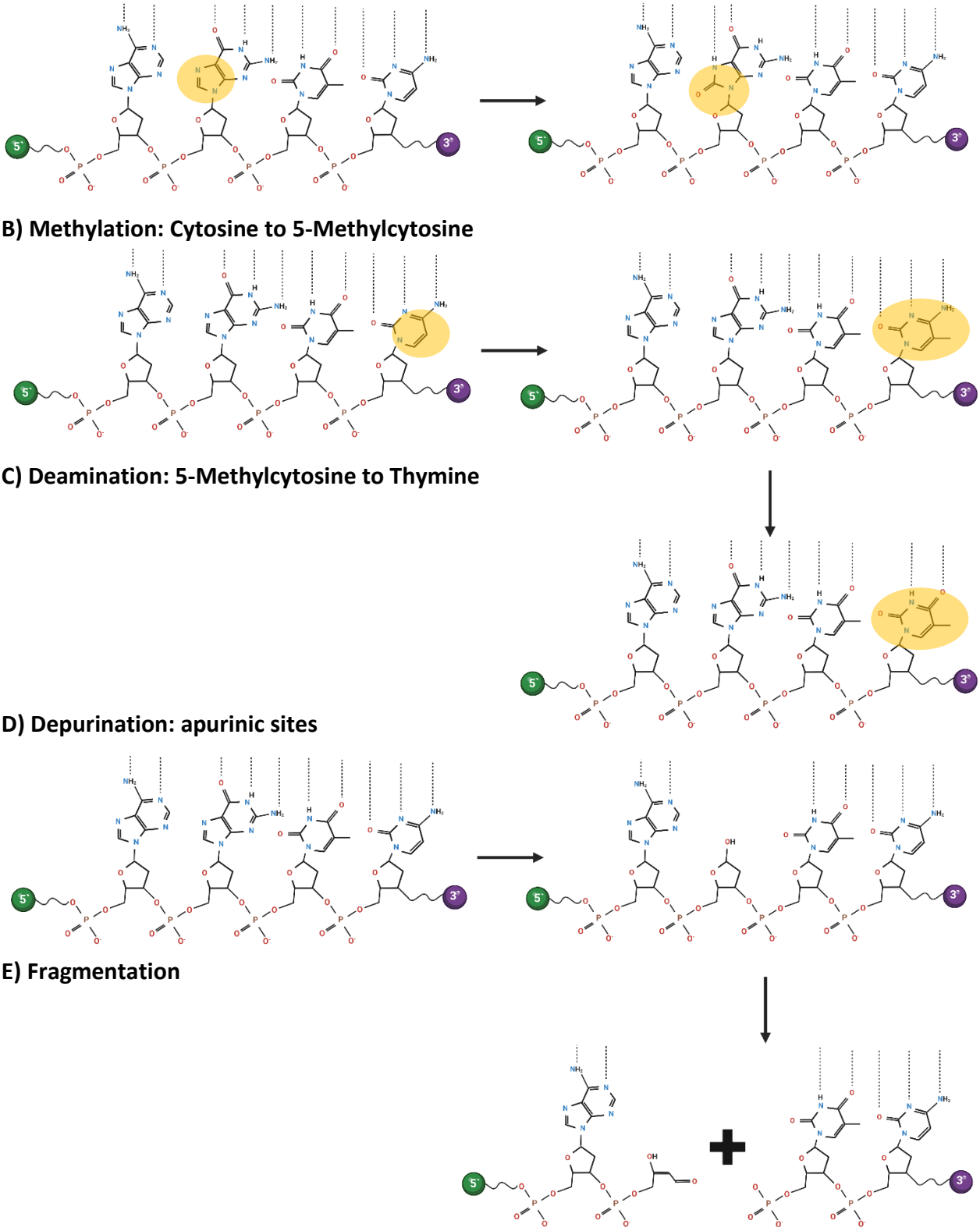


Figure 5: Common chemical modifications of nucleic acids.

the identification of so-called CpG motifs, which are unmethylated cytosine-guanosine dinucleotides contained inside a certain pattern of surrounding bases. [37] A recent study has shown that DNA packaged in adeno associated viruses has little to no methylation in the CpG motifs [38] and thus leads to immune activation. [39] These examples demonstrate that DNA methylation needs to be closely monitored and fine-tuned on a case-to-case basis to balance transcription repression and immunogenicity.

Deamination of (methyl) cytosine is another commonly described reaction of NAs. It has the potential to transform cytosine to uracil and methylcytosine to thymine and thus significantly alter the genetic information. The spontaneous reaction is reported to be very slow for single stranded NAs with a half-time of approx. 200 years and ca. 140 times slower for double stranded DNA at 37 °C at pH 7.4. [40] However, certain conditions like high pH or temperature significantly increase the reaction speed so this variant might be of concern in some NA therapeutics. Depurination and depyrimidination are both spontaneous hydrolytic reactions which bring the risk of mutagenesis and represent the most abundant endogenic DNA damage. [33] The reaction is generally faster for purines and single strands and favors acidic conditions with a half time of around 200 h at pH 2.5. [41] It is one major limitation in oligonucleotide synthesis, since in the solid state synthesis an acidic reaction step is necessary. As an additional concern, the resulting free ribose sugar is very reactive and may lead to fragmentation. A study from Calabretta and colleagues found that the effect of oxidation and abasic sites is strongly impacting ribosomal translation of mRNA. [42] While oxidation was classified as a less severe lesion that leads to a mixture of truncated and full-length polypeptides, abasic sites interrupted the translation process in total. As with protein biopharmaceuticals small chemical changes can impact the overall structure of the whole NA and thus lead to reduced safety and efficacy of the drug product.

1.3 Selection of analytical test methods to cover quality attributes

The simplest technique to verify certain critical quality attributes (CQAs), such as the purity and appearance of pharmaceutical items, is through visual inspection. This includes, for example, deviation from the expected color of the product but also the detection of particles in the product that may be caused by impurities, aggregation or suspended particles. The integrity of the packaging and fill level of the product can also be inspected in this way. Possible damage, such as cracks, breaks or leaks, can affect the sterility or stability of the product. [43] Some general chemical properties that are the underlying of CQAs can be determined with abundant laboratory equipment such as pH meters or UV-photometers. Yet, for the vast majority of quality attributes, specialist equipment is required. Analytical methods can be classified based on their separation

principles (see Figure 6) and this chapter aims to give an overview on the analytical toolbox that is used to address the previously described CQAs.

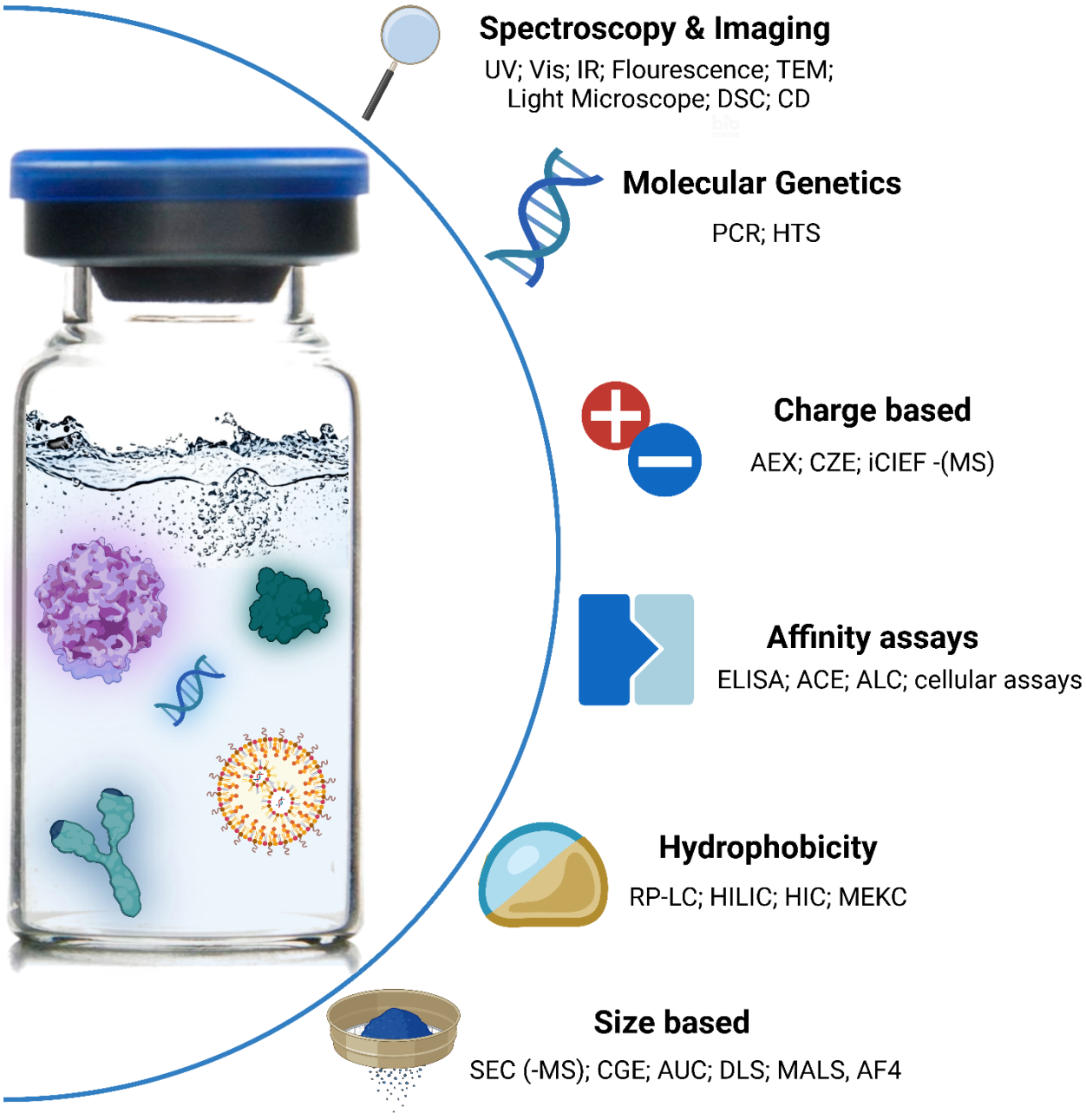


Figure 6: Classification of prominent analytical techniques based on their separation principle

1.3.1 Spectroscopic techniques

Spectroscopy refers to a class of physical methods that monitor the interaction between electromagnetic radiation and matter. The range of electromagnetic radiation wavelengths that each individual spectroscopic technique can use, serves as the primary defining characteristic between various approaches. [44] Generally, these interactions occur through the initial absorption of a portion of the incoming electromagnetic radiation. Fluorescence allows for the reemission of this absorbed electromagnetic radiation at different wavelengths which is detected by spectroscopy. Even though spectroscopy is a large field that encompasses numerous methods, biopharmaceutical scientists are typically interested in techniques that make measurements in the ultraviolet (UV), visible (VIS), infrared (IR), and radio wave regions of the electromagnetic spectrum. [44] These spectroscopic techniques have the capacity to offer details regarding the physicochemical surroundings of each of the functional groups that engage in electromagnetic radiation interaction. Because the physicochemical environment of these groups is likewise affected by the biopharmaceutical's folding state, spectroscopic approaches can offer indirect information when the structure of a biopharmaceutical changes. UV/VIS and fluorescence spectroscopy use electromagnetic radiation in the UV/VIS range (190-700 nm), whereas Fourier-transform IR (FTIR) spectroscopy uses electromagnetic radiation in the range of 0.7 µm to 1000 µm with 2.5 µm to 15 µm being the most frequently used range. [45]

UV/VIS spectroscopy is utilized in biopharmaceutical development for a variety of purposes. On the one hand, it may be used to characterize proteins and NAs. To this end, characteristic absorption bands are monitored, allowing judgments about the composition and structure of the molecules to be derived. One example would be the hypochromic effect, which can be used to distinguish between single-stranded and double-stranded DNA and thus determine melting temperatures by measuring absorbance at 260 nm at varying temperatures. [46] Purification procedures can also be monitored. The quotient of absorbance at 260 and 280 nm, for example, can be used to make inferences about the purity of the protein or NA and confirm that there are no contaminants. [47] Nevertheless, the most abundant use is the quantification of API concentrations. [48] The most precise approach uses a direct absorbance measurement for the intrinsic chromophores in the UV-range, but colorimetric variations like the Bradford protein assay may be useful in case of nucleic acid contaminations or proteins with no aromatic residues. [49] Given the simplicity, speed and reliability of UV/VIS spectroscopy, it is performed almost exclusively to determine the drug concentration. Moreover, UV/Vis spectroscopy serves as the foundation for detectors used in chromatographic and electrophoretic techniques, allowing the

separation and quantification of substances in a sample based on their absorption, e.g. at 280 nm for mAbs or at 260 nm for NAs.

When light of a specific wavelength is absorbed during fluorescence emission spectroscopy, an electron moves from its ground state to its excited state. Energy is lost as the electron returns to its ground state, and the system emits light with a longer wavelength and lower energy. Fluorescent molecules typically have aromatic groups, which in the case of biopharmaceuticals include aromatic amino acids and nucleobases. [45] In addition, the physicochemical behavior of the fluorophore is extremely sensitive to its physicochemical environment. In general, aromatic amino acids, due to their higher hydrophobicity, exhibit a propensity to be integrated into the central region of a protein, thereby reducing their exposure to the surrounding solvent. It is commonly observed that the intrinsic fluorescence of a protein tends to exhibit a red-shift during the process of unfolding. However, it is also possible to observe a blue-shift in some cases. [50] For instance, a deeply embedded tryptophan exhibits an emission peak at approximately 330 nm, whereas a tryptophan that is prominently exposed to the surrounding solvent on the protein's surface, displays an emission peak in the range of 340-350 nm. The process of protein unfolding, such as through the use of a denaturant, leads to an increased exposure of these residues to the surrounding solvent. [50] As for UV/Vis spectroscopy, the use as detection system for chromatographic or electrophoretic systems is of utmost importance. [51] Many biomedical substances have intrinsic fluorescence but also reactive groups can be used to couple strong fluorophores. For DNA intercalating dyes are widely employed for this purpose, [52] while for proteins typically covalent reactions e.g. at a Lysine residue are performed. [53] Derivatization improves analytical sensitivity significantly when compared to UV/Vis detection but may add a bias due to selective accessibility of the used dye to the analyte. [54]

The absorption of electromagnetic radiation in FTIR spectroscopy is related to the vibrational properties of the amide bond in the polypeptide backbone. By this, the amide groups in the polypeptide chain of the protein are assessed for their flexibility. [55] FTIR analysis is a useful tool for analyzing changes in a protein's secondary structure because the peptide bond is part of folding interactions involved in protein stabilization. The technique is thus employed in stability characterization studies [56] as well as in process controls during manufacturing [57]. However, the hyphenation of FTIR with CE and LC is difficult, though technically possible, due to absorption of water that is present in the mobile phase or background electrolyte. [58]

1.3.2 Hydrodynamic techniques

Methods that study the hydrodynamic properties of biopharmaceuticals are mainly used to investigate the quaternary structure and thus aggregation. Global properties are studied by monitoring the movement or transport of the molecule through a fluid medium in response to a driving force. [44]

In dynamic light scattering (DLS), this force is the thermal kinetic energy of an analyte that leads to Brownian motion. Light is scattered in all directions when it encounters tiny particles or macromolecules in suspension or solution (Rayleigh scattering). When coherent, monochromatic laser light is employed, interference causes slight variations in the scattering intensity because Brownian molecular motion constantly alters the distances between the scattering sites. [59] By recording these fluctuations on a time scale, information is thus obtained about the speed at which the particles move in solution. This can be used, in turn, to calculate a diffusion coefficient, from which the hydrodynamic radius can be determined according to the Stokes-Einstein relationship. [60] DLS thus provides size distribution plots which is particularly useful for the control of aggregation and particle formation. In the majority of contemporary instruments, the detection process can occur simultaneously at multiple angles through either an array of detectors positioned at predetermined angular locations or a single photodetector capable of adjusting its angle. This method is thus called multi-angle light scattering (MALS). [61] A major advantage of MALS is that it is independent of the sample, meaning it can be used without prior knowledge of the sample or the specific molecular structure. MALS can be used with a variety of biopharmaceuticals, such as proteins and antibodies (including chemically modifications like PEGylation) NAs, and liposomes, even up to nanoparticles, because it does not require for labeling or sample modification for this testing. Another advantage of MALS is that it provides the ability to obtain additional information about the composition and conformation of the sample. [59] By combining MALS with other techniques such as Size Exclusion Chromatography (SEC) or other chromatographic methods, information on the relative composition and conformation of biomolecules can be obtained. [62] However, the DLS and MALS technique also have some disadvantages. One point is the limited resolution as related molecules (such as monomers and dimers) cannot be resolved in batch mode. Additionally, the technique is prone to error and artifacts if not handled by an expert analyst as it can be affected by impurities, turbidity, streaks or air bubbles. [63] In modern devices, results are displayed after complex software processing, which can be a source of ambiguity when the fitting process is not closely monitored.

Analytical ultracentrifugation (AUC) is a technique that integrates an ultracentrifuge with optical monitoring systems. By utilizing specialized instrumentation and adjusting measurement variables, various types of experiments can be conducted. The most frequently conducted AUC experiments involve sedimentation velocity (SV) and sedimentation equilibrium. Among these, the SV mode is considered the most significant. In SV-AUC the movement of the analyte is induced by a strong centrifugal field (typically above 100 000 g) and tracked by an UV, interference or fluorescence detector. [64] Such experiments reveal information about a broad range of biopharmaceuticals regarding size and shape, sample purity and oligomerization or aggregation states. [65] In most cases the sample can remain in its native environment, potentially even regarding concentration, and no sample preparation might be required. By adjusting the centrifugation speed a broad size range from single kilodaltons to gigadaltons can be covered. [66]. While this technique was historically the gold standard for biomolecule analysis, today it is mainly used for characterization purposes and as an orthogonal approach to confirm results from other approaches. This is mainly because of relatively large sample consumption (>100 uL), low throughput and high labor and data processing requirements. [44] However, in addition to purity analysis, AUC is also one of the few techniques that can be used to determine the content ratio of viral vectors which is part of the “quantity” CQAs (Figure 2). For a variety of recombinant viruses like the rAAV a majority of produced viral particles does not contain the genome and thus the ratio of empty/genome containing particles has to be determined as quality measure. [67]

1.3.3 Imaging techniques

Parenteral medications may contain intrinsic, extrinsic and inherent particles from a variety of sources, including fibers, glass, rubber, metal, or silicone oil droplets produced by packing and storage materials. [68] Yet, due to the API's tendency to aggregate, especially with increasing concentration in solution, the biophysical characteristics of a protein biopharmaceutical cause a special inherent problem of particle formation. The additional source of particle formation from the API itself and the accelerating effect of foreign material, which may enhance particle formation, exacerbates the property of protein-based biopharmaceuticals to cause immunogenic effects. [69] As a result, for biopharmaceuticals, the existence of particles is considerably more important than for small molecule pharmaceuticals. Particles are very diverse and can vary in size over six magnitudes, from nanometers to millimeters. [70] Visible particles (>100 µM) are often assessed through direct physical inspection. [71] For subvisible particles, a classical light microscope that uses the visible light spectrum is able to visualize particles larger than 400 nm. This approach may be used for particle quantification, but is rather low throughput, gives little information on the kind of particles and results are dependent on user training. [72] For this

reason, subvisible particles are typically detected by particle counters. These devices have a flow cell with controlled media flow and use light scattering detectors for particles smaller than 1 µm or light blocking/obscuration for larger particles. [73] The light obscuration particle represents a standard method for the analysis of clinical and commercial injectable pharmaceuticals. This test is founded on the principle of light blockage and enables an automated determination of particle size and the quantity of particles based on their respective sizes. [71] More modern instruments combine a microscope with a flow cell and are thus able to directly image particles with a high definition, high magnification camera. Subsequent data processing can help distinguish different classes of particles and diminish user variability. [74]

Transmission electron microscopy (TEM) visualizes objects by using electrons rather than light. The electron beam is sent through a vacuum tube onto the target being studied. It is refracted and dispersed as it travels through the item, and thus carries information about the structure of the sample. Specimens must be fixated during this process to generate a clear image. The electron beam is then captured by a detector and converted into an electrical signal, resulting in a detailed image. [73] Electron microscopes (EM) can visualize objects with nanometer resolution, through the significantly lower wavelength of electrons compared to visual light. Biological samples may be cooled down very quickly in order to achieve fixation (Cryo-EM). This approach brings the benefit, that the sample can be imaged in its actual environment. Cryo-EM is an established technique for structure elucidation of biomacromolecules that reaches a resolution in the Ångström region. This is achieved by combining information of thousands of pictures with short exposure times of a molecule in a statistical model. [75] The problem of limited exposure time can be circumvented by negative staining. The analyzed objects are embedded in an amorphous substance such as uranyl acetate with a high scattering power in this procedure. Consequently, images of objects with great contrast can be obtained even when they are hardly apparent in standard preparations. [76]. This method has been demonstrated to provide detailed information on aggregate microstructure and to provide insights into the conformational integrity of the component proteins. [77] Despite its unique features and advantages regarding information delivery on molecular level, TEM suffers significant drawbacks. The instrumentation is expensive, the technique is low-throughput [78] and even though automation is possible, quantification remains challenging. [79] For this reason, it remains a niche method for late-stage protein biopharmaceutical characterization. A drawback of Cryo-EM technique is, that the strong energy of the electron beam damages the sample over time limiting the contrast on a specific molecule. [80] These drawbacks of low-throughput and high instrumentation cost thus limit routine use of this technique in practice. Historically, electron microscopy was a technique that was widely

employed to study a certain class of particles that could not be detected with light microscopes: viruses. With the emergence of viral vectors for C&GT the unique feature of providing direct visual information of size, shape, and aggregation of the vectors once again proofs to be of high value. The very first description of negative staining in 1959 mentions “ghost” particles of turnip yellow mosaic virus which do not contain NAs. [76] These empty viral capsids are now considered a CQA in modern gene therapy products. Consequently, negative stain TEM has been described as the gold standard for the determination of the full/empty ratio [81] and there are attempts to offer a fully automated data interpretation without tedious counting. The technique is also employed alongside Cryo-EM to monitor morphology of LNPs. [82,83] However, the contrast can be difficult to interpret for partially filled capsids and no information can be gained on wrongly filled particles. [84]

1.3.4 Mass spectrometry techniques

Mass spectrometers operate by transferring molecules from a gas or condensed phase into a vacuum, ionizing them, and then transporting them with the help of electromagnetic fields. This allows to evaluate the abundance of ionic species based on their mass to charge ratio (m/z). [85] The most frequent approach for the generation of biomolecular ions from solutions is electrospray ionization (ESI). The analyte in solution is brought to the ESI source via a thin capillary, making this ionization approach easy to connect with liquid chromatography (LC) or capillary electrophoresis (CE). [86] When a strong electrostatic potential is applied to the tip of the capillary, a thin jet of charged liquid is created. Dispersion of this jet generates charged aerosol particles, which evaporate a significant amount of solvent and thus raise the droplet surface charge over the so-called Rayleigh limit. [85] Whenever the droplets have reached the Rayleigh limit, they begin to disintegrate, which results in the formation of smaller charged droplets. During numerous iterations of this process on increasingly smaller levels, the analyte molecules are finally released from the droplets. This results in the generation of ions that, depending on their dimensions, may hold several charges simultaneously. [86] The capacity of mass spectrometers to categorize ions according to their m/z values is founded on the distinctive impact of electric and magnetic fields on the mobility of ions in vacuum in a mass analyzer. In practice, this is accomplished by using a mix of electrostatic and electrodynamic fields (ion traps and quadrupoles), electrostatic fields (time of flight /TOF MS), or static magnetic field (ion cyclotron resonance /ICR MS). [86] As an ion passes by or strikes a surface, a detector records either the charge generated or the current created. [85] The examination of biomacromolecules may be conducted either from the bottom up or the top down. Bottom up refers to the digestion

of the analyte with enzymes and subsequent analysis of the peptides or oligonucleotides, while the top down technique corresponds to the analysis of intact biopolymers. [87]

For native polymers, the average molecular mass is calculated in most cases. However, if the mass analyzer has adequate resolution, individual macromolecule isotopes may be distinguished from one another. Isotope resolution of smaller molecules up to 15 kDa is possible with TOF mass analyzers [88], while isotope resolution of bigger constructs (<150 kDa) was demonstrated with Fourier-transform ICR mass analyzers [89]. Since the accurate monoisotopic mass of bigger molecules is actually not determinable and the molecular envelope width is rather wide, the isotopic resolution of larger molecules provides only a modest amount of extra information above the average molecular mass, making the benefit of isotopic resolution for larger molecules debatable. So far only limited approaches have been presented for long NAs and the top-down analysis is mostly limited to proteins. [90,91] As a consequence of the phenomenon of multiple charging that occurs under ESI conditions, even the largest proteins are able to fall within the mass range of widely used mass analyzers, which is from m/z 500 to 4500. [86] The information gathered of analytes that are composed of numerous polypeptide chains, such as mAbs, may improve through preparation of denatured and reduced samples. This combination of tests offers an initial validation of the gene-derived protein sequence, indicates the structural integrity, and further discloses post-translational modifications. [92]

The measurement of intact proteins does often not provide the full picture. For example, it may not reveal the complete amino acid sequence and has limitations in localizing PTMs. Through a proteolytic digestion by, e.g. trypsin or Lys-C, further structural details are revealed at the peptide level. [86] This method is known as bottom-up analysis or peptide mapping and often employs multi-dimensional-MS. A multi-dimensional mass spectrometer is a device capable of successive cycles of mass spectrometry, typically including some type of molecule fragmentation. [85] For instance, a mass analyzer can separate a single peptide from the numerous ones that enter a mass spectrometer. Another second mass analyzer then stabilizes the peptide ions when they collide with a gas, triggering their collision-induced dissociation (CID). The fragments formed by the peptides are then sorted by a third mass analyzer. [86] As the cleavage to peptides dramatically increases the sample's complexity, a high quality separation e.g. by reversed phase LC or capillary electrophoresis [93] is necessary before MS analysis. Through a peptide map it is possible to retrieve information on nearly all PTMs, including glycosylation, phosphorylation, and oxidation can be detected including their position and composition. [94] In addition, it is considered the safest approach for identity testing of proteins. Bottom-up approaches are also applied for NA therapeutics. NAs are only formed by four nucleotides, in contrast to proteins,

which are built out of 20 amino acids. This provides the issue of generating isobaric product ions that cannot be differentiated by MS. [95] In addition, the strong salt adduction that is often seen for NAs may sometimes produce ion suppression and make spectra interpretation demanding. Still, it remains a highly useful approach to not only detect modifications such as deamination or methylation but also quantify them. [96] Through its ability to analyze multiple CQAs at the same time it increasingly gained popularity over time and it is now also referred to as multi attribute monitoring. Today, it is inconceivable to submit dossiers to the regulatory authorities without including comprehensive MS characterization data. [97] Despite this, MS does have a few downsides. PTMs might be lost during digestion or CID. Hence, these methods carry with them the risk of losing essential information that connect with molecule function. A problem that is not directly caused by MS but by enzyme incompatibility is posed because of the increased stability of changed residues. This could limit the digestion efficiency. [98] MS devices remain relatively expensive and the large amount of produced data can complicate the analysis.

1.3.5 Bio(-chemical) assays

During biotherapeutic development, it is necessary to conduct a bioassay that is both selective and physiologically relevant in order to provide data on the product's potency and stability as well as reflect the intended mode of action. Bioassays can be costly to design, conduct, transfer, and maintain, which complicates the selection of a suitable technique. [99] In spite of attempts to adopt control measures, cell-based bioassays are naturally imprecise and often miss the accuracy and robustness of biophysical approaches due to their reliance on live organisms, tissues, or cells. The use of a strategy for the stepwise creation of bioassays is typically supported by regulators. In the initial stages of product development, it is often advised to begin with a binding approach, such as enzyme-linked immunosorbent assays (ELISAs) for easier implementation. Still the immobilization of the target and the production of a suitable detection antibody remains challenging. In subsequent phases, this strategy enables designing more complicated cell-based bioassays. [99]

The most prevalent kind of biochemical tests are binding assays, like ELISA. It relies on the specific interaction between an antibody and an antigen and is thus limited to protein analytics. In the reaction, one of the agents has been tagged with an enzyme which causes the production of a colorimetric product, enabling spectrophotometric detection. [85] As a consequence, the signal intensity is proportional to the concentration of particular antibody and antigen. On a multiwell plate, the reaction may be configured in numerous ways, with the sandwich assay serving as the most common format for quantifying a target antigen or analyte. [100] During research,

production, and manufacturing, ELISA is a powerful tool for the quantification of the desired protein product. It is a helpful instrument for selective quantification of a desired protein from complex protein mixtures such as cell culture fermentation media or product pools in initial phases of protein extraction as well as downstream processing. This is because ELISA is highly selective and has mostly a low susceptibility towards matrix effects, i.e. even complex mixes do not need considerable sample preparation. [100] Nevertheless, it is vital to verify that the sample matrix is free of any buffering agents, salts, surfactants, or protein molecules that might disrupt the antibody–antigen reactions of the ELISA. Another application involves the quantification of viral capsid titer in gene therapy products. [101] In addition to quantification, ELISA is an effective approach for determining the presence of certain contaminants, such as aggregates [102] or host cell proteins [103]. Furthermore, ELISA may be used to quantify the binding that occurs between a therapeutic mAb and its particular target antigen, and as a result, it can serve as a straightforward activity test. [104] As most biochemical methods, ELISA suffers from its imprecision which is typically 20% relative standard deviation but can be as high as 30%. In addition, the development of specific antigens/antibodies can be difficult and the dynamic range is rather narrow. [100]

The binding of a therapeutic mAb is only involved in the first phase of a biological event, which is often followed by numerous reactions like the antibody dependent cytotoxicity or complement dependent cytotoxicity. [105] An ELISA is not capable of providing a complete representation of the mode of action of biopharmaceuticals in most cases. Detection and quantification of a medication's mode of action and implied higher order structure, potency and efficacy are distinctive to cell-based bioassays and, as such, they are an anticipated component of any analytical bundle for a biotherapeutic. [99] Bioassays, unlike other analytical procedures, are typically exclusively designed individually for each combination pharmaceutical and targeted mode of action. Thus such assays must have a great variety depending on the detected cellular mechanism and respective readouts. Bioassays may assess cell proliferation, death, cytotoxicity, or the production of cytokines or membrane markers. [106] The very first efforts to assess biological activity consisted of in vivo bioassays, in which protein was delivered to animals and their responses were evaluated. As a result of attempts to limit animal testing, the preferred procedure nowadays are in vitro tests, in which cells or tissues are cultivated in the laboratory and utilized as responders to the test substance. [107] Typically, a cell-based bioassay process involves cellular proliferation in suspension or adherent cultures over multiple days/weeks. In a next step, various concentrations of the therapeutic of interest and a reference standard are added in order to activate or inhibit cellular processes in a multiwell assay format. The end-point

response may be noticed anywhere from several hours to several days later, depending on the mechanism of action. There is a huge array of readouts and detection methods that may be used since readouts for an experiment may be obtained from any stage of a cell activation cascade. [107] As an example polymerase chain reaction may be used to detect mRNA transcription or flowcytometry for protein expression. The selected readout, which represents the magnitude of the biological reaction to each concentration of the therapeutic, can be fitted into an equation for data analysis. [107] Assuming that a cell-based bioassay may be representative for the mode of action and labor- and cost-efficient, they still often need a lengthy period of operation and exhibit relatively significant assay variability, with cellular effects typically requiring days to manifest. [99] For this reason, there are attempts to simplify this crucial testing e.g. through reporter gene assays. [106]

1.3.6 Molecular biology techniques for nucleic acids

Clinicians and researchers rely heavily on polymerase chain reaction (PCR)-based methodologies to diagnose illnesses, clone and sequence genes, and conduct advanced quantitative and genomic analyses with great sensitivity. [108] Recently, the technique became well known in the public again through its medical application for detection of SARS-CoV-2 pathogens. [109] The method is used to generate hundreds to millions of copies of a specific DNA sequence by amplifying a single or a few copies of a piece of DNA by many orders of magnitude. The PCR is based on a three-step cycling process: first the denaturation of double-stranded DNA, next the annealing of primers, and last the extension of primers by the polymerase. [110] Denaturation splits the complementary DNA strands kept together by hydrogen bonds in a duplex structure. Heating the sample to 95-100 °C is a straightforward and effective method to accomplish this. Primers (typically 15 nt ssRNA) are bonded to the separated DNA strands during annealing at ca. 55 °C. [108] Each primer is complementary to either the 5' or right 3' end of the target sequence. After annealing, thermostable DNA polymerase catalyzes the synthesis of complementary DNA strands onto the annealed primer at ca. 70 °C. [111] This process is automated in a thermocycler. With completion of one PCR cycle, the amount of DNA strands doubles. After up to thirty cycles, a single molecule of target DNA may be amplified to approx. one billion copies. For a simple analysis and confirmation of the PCR product qualitative detection may be sufficient, for instance using agarose gel electrophoresis. [110] PCR is not limited to DNA: By reverse transcribing RNA to complementary DNA quantification and identification of RNA is feasible, too. Quantitative or real-time PCR (qPCR, RT-PCR) displays the amount of a DNA in a reaction chamber after each replication cycle. [111] Thus, qPCR permits simultaneous detection and quantification of the PCR product as it is being generated. Most commonly, fluorescent dyes which non-specifically

intercalate with dsDNA are employed. [112] Digital droplet PCR (ddPCR) is a variant of PCR in which the PCR reaction mixture is initially divided into nanoliter droplets containing zero, one, or multiple template DNAs. After the reaction cycles, individual droplets are studied based on the fluorescence intensity of a given droplet. [113,114] Therefore, it is possible to identify whether one or multiple target molecules were present in the droplet prior to the initiation of heat cycles. The method has the benefit of direct and self-reliant measurement of DNA without standard curves, resulting in more exact, repeatable and sensitive results compared to qPCR. [113]

In the biopharma industry, qPCR is a routine test technique for quantification of DNA. For protein formulations, it can be employed to find traces of host cell DNA impurities. [115] PCR is the gold standard for the identification and quantification of NAs in therapeutics comprising longer NAs like viral vectors or LNPs. [116-118] It remains unrivaled in terms of sensitivity, since, theoretically, it can identify a single target molecule. On the other hand, the possibility of cross-contamination with amplifiable DNA may make this high sensitivity problematic. [118] In addition, only a tiny portion of the NA gets detected in practice. When just one set of primers is employed, this may lead to inconsistent findings.

In 2003 the human genome project was declared being complete. This enormous development included sequencing all of the roughly three billion base pairs of DNA that comprise the human genome and is considered the largest single undertaking in the history of biological science. [119] The project took thirteen years to complete costing more than five billion dollars. [120] Since then, remarkable advances in DNA sequencing technology have been accomplished. Nowadays, a human genome can be sequenced for a few thousand dollars with a one-day turnaround time. [121] While the term “next generation” sequencing (NGS) is frequently used it may be misleading as today the literature describes up to four generations of different techniques. [122] For the sake of clarity the term, high throughput sequencing (HTS) will be used. One of the most widely used methods is called Illumina sequencing and is often counted as part of the second generation sequencing. In this technique, the genetic material is fragmented randomly into approx. 300-500 bp reads. Each end is ligated with an appropriate adaptor and utilized as a substrate for the repeated amplification synthesis process. This sample preparation is also referred to as library generation. [123] A glass slide containing oligonucleotide sequences corresponding to a ligated adaptor is used to carry out the sequencing. Oligonucleotides on the slide are arranged in a way that successive cycles of DNA amplification result in the formation of clonal "clusters" containing about 1000 copies of a fragment. [122] When putting the library into the flow-cell, the dsDNA fragments are denatured into ssDNA molecules. As soon as they are placed on the flow cell, they hybridize to the oligonucleotides that are located on the surface of the chamber. These oligo

nucleotides serve as primers and are utilized to produce an initial copy of the specific DNA fragment. [124] The original library fragments are discarded, and their copies that are connected to the flow cell are used in an isothermal amplification process in order to build a cluster of equivalent template molecules. At 60 °C, this PCR is accomplished by cyclic alternations of three specialized buffers that facilitate the denaturation, annealing, and extension phases. [124] Throughout these operations, the 3’ end of the replicated library molecules have the chance to hybridize with the complementary oligonucleotides on the flow cell, therefore producing a bridge structure. For this reason this technique is often referred to as bridge amplification. [122] Using the cleaving site in the surface oligo, one strand of the dsDNA fragments is removed in the last step, releasing the “bridge”. Bridge amplification is an inefficient technique for clonal amplification, since 35 amplification cycles only generate 1,000 copies of the original molecule. [124] Nevertheless, every glass slide can accommodate millions of cluster reactions simultaneously, which is the foundation of the speed of this approach. Across the course of the synthesis steps, proprietary modified nucleotides are inserted into the growing DNA chain. These nucleotides correspond to all four bases and are each tagged with a distinct fluorescent label which serves for detection as soon the base is incorporated. [122] In addition to this, the nucleotides carry out the role of synthesis terminators for each reaction, which, once detected, are released to allow the subsequent cycle of synthesis to take place. At least 300 rounds of the elongation reactions are performed, with one round lasting 10-60 min depending on the device. [124] The vast information gathered from millions of clusters is matched through overlaps in the first readings and sorted into the entire DNA sequence by sophisticated data processing systems. Sequencing is a very powerful technique for the development of biopharmaceuticals, especially in the early research. It can be used in nearly all fields such as target identification, biomarker discovery or study of genetic variations. [125] Moreover, HTS is a fundamental component of personalized medicine, which takes into consideration the patient's surroundings and habits, medication history, and their entire genome. [126] It has been reported to rapidly and precisely filter host sequences and categorize viral readings for the identification of host DNA contaminants and the protection against adventitious agent safety. [127,128] This aspect is of special interest for viral therapeutics, since the DS is a virus itself and reversion mutants are a major concern. Through its holistic approach it is a good choice to verify identity of NA therapeutics with longer sequences, which cannot be targeted with MS, and can also track down unknown impurities. [129] HTS remains expensive in routine use due its complexity and the need for specialized staff. [118,130] Moreover, the library preparation is complicated and can add biases and the steadily changing sequencing instrumentation makes the definition of standards for QC use difficult.

Considering the rapid advancements of sequencing during the past twenty years, it can be assumed, that the importance of HTS will steadily increase through innovation and cost reduction.

1.3.6 Chromatographic methods

Most liquid-chromatography techniques exploit direct enthalpy interactions that engage bonding interactions between the chemical groups on the surface of a biopharmaceutical and the stationary phase of the chromatographic material under certain solution conditions. [44] Historically this chromatography was performed at atmospheric pressure in glass columns. However, modern devices employ high pressure (>3 MPa) and thus carry the name high pressure/performance liquid chromatography (HPLC). [85] One type of chromatography which is highly relevant to the biopharmaceutical industry is not reliant on enthalpic interactions: size exclusion chromatography (SEC). As a chromatographic technique, SEC separates biomacromolecules based on their hydrodynamic volume in solution. Separation, like in other chromatographic modes, is accomplished by a partitioning process. In SEC, the stationary phase is the liquid located within the pores of chromatographic particles and it shares the same chemical composition as the mobile phase. [85] The access to the stationary phase is carefully controlled by the pore structure of the chromatographic beads or particles. In contrast to other modes, SEC's separation process is based on entropic factors and the presence of enthalpic interactions is considered deleterious. [131] A large molecule that cannot enter any of the smaller pores can only enter the interparticle volume and elutes early after this volume of mobile phase has migrated through the column. On the other hand, if a small particle is able to penetrate each part of the stationary phase pore system, it has to pass a volume equal to the sum of the pore volume and the interparticle volume. This small molecule elutes after it has permeated approximately double the volume of the large molecule and thus will be detected at the end of a run. [132] The retention mechanism provides general benefits to this technique. The entire sample is eluted within one column volume, based on its hydrodynamic size, and (optimally) no analyte is remaining. [133] Taking into account suitable mobile phase conditions, there is low potential for conformational alterations since the separated molecules should not interfere with a surface, e.g. the stationary phase. For this reason, this approach is exceedingly mild and protects the function of biomolecules very well. [85] However, since only substances with molecular sizes between two rather narrow thresholds may be separated on a given column, SEC has a modest peak capacity. As a result, it is a poor resolution method, giving at most 6-8 peaks and needing a 1.5 to 2 fold difference in molecular weight for resolution. [133] SEC's biggest applicability in the industry is its capacity to analyze the degree of aggregation i.e., high molecular weight species, HMW. It can typically monitor aggregation with an estimated limit of quantification (LOQ) of 0.1% to 0.5% total

aggregation, whereas AUC is in the 2% range. [134] While SEC represents the standard method for HMW analysis, the nature of these apparent aggregates should be interpreted with care. Possible causes of SEC data misinterpretation include inherent fragment heterogeneity, interactions between the analyte and/or its variant forms with the chromatographic material, and HMW contaminants from excipients. Moreover, HMWs can be diluted during the injection process and can then dissociate during separation. [44]

Apart from SEC, chromatography is a technology for investigating a biopharmaceutical's unique surface chemical topology in terms of its interactions with the chromatographic surface in a specific solution. [44] Normal phase (NP) chromatography is commonly used for adsorption chromatographic separations where the stationary phase is more polar than the mobile phase. Silica gels with spherical or irregular shapes feature free silanol groups and adsorptive characteristics due to aluminum and oxygen ions on their surface. [85] Sample molecules with free electron pairs, functional groups, or double bonds can form hydrogen bonds, dipole-dipole interactions, induced dipoles, and complex bonds with these adsorbents. Adsorption columns are typically eluted with strongly nonpolar liquids (hexane, isooctane). [133] Because these compounds have considerable impacts on biopolymers, NP columns are rarely employed except in hydrocarbon analysis. [85] The surface silanol groups of silica gels can be chemically altered with different polar or hydrophobic functional groups to vary the interaction mode. The reverse phase (RP-)HPLC stationary phase is created by reacting silanes that have been substituted with long-chain hydrocarbons with silica gel. Thus, the polar surface of the silica gel particles has been covered with a non-polar layer of alkanes, “reversing” the polarity. [135] In the characterization of biopharmaceuticals, RP-LC is used at various levels for protein biopharmaceuticals with mobile phases that often mix water or buffer and acetonitrile with e.g. trifluoroacetic acid (TFA) as typical additive. Separation relies on solvophobic and electrostatic interactions, the latter being directed by the pH of the mobile phase and ion pairing interactions of TFA with basic side chains. [85] Changing the polarity of the mobile phase by increasing concentrations of organic solvent thus leads to elution of the analyte. Temperature is frequently utilized as an additional variable to cope with the slow diffusion of proteins in liquids, which may result in large peaks. The method is able to detect various PTMs such as Met-oxidation, fragmentation, deamidation and glycosylation. [92] RP-LC outperforms most chromatographic modes in terms of efficiency and its robustness makes it suitable for usage in a routine environment. It is the most coupled mode with MS, which makes it essential for characterization purposes. [86] The main limitation is the use of relatively harsh parameters (organic solvents; high temperature and acidic conditions) which not only denature the protein but can also induce fragmentation/clipping as artifacts. [136]

Ion pair (IP)-RP-HPLC is frequently employed to separate NAs by size. By adding counterions to the mobile phase the analyte forms ion pairs with them. Resulting ion pairs are neutrally charged and are separated by hydrophobic interactions with the reversed phase and liquid-liquid chromatographic partitioning processes. [85] Furthermore, the hydrophobicity of the bases in the sequence is a second important element that impacts separation. [137] Triethylamine (TEA) forms a cation in acidic mobile phases and is commonly utilized as a DNA counterion. As the number of phosphodiesters increases with NA length, more counterions bind, resulting in longer molecules eluting later. [138] The technique is widely used for purity analysis of short, synthetic oligonucleotides, as it can also achieve a certain degree of sequence specific resolution between two NAs of the same length. [139] Nevertheless, resolution decreases rapidly beyond 1000 nucleotides, which is one of the main limitations of this approach.

Hydrophobic interactions with the stationary phase describe the core separation process mechanism in both, RP-LC and hydrophobic interaction chromatography (HIC). The main difference in HIC is, stationary phases with a hydrophobicity that is only approximately 10% that of RP materials. Wide-pore hydrophilic gel filtration materials bonded with weakly hydrophobic alkyl groups being most commonly utilized for this purpose. [85] Through high salt concentrations in the mobile phase, which improve protein hydrophobicity, a tight contact between stationary phase and the area of the protein involved in adsorption can be facilitated. Yet, the salt not only changes the conformational structure, but it also affects the molar concentration of water. As a result, water molecules are the displacing agents in this process, and proteins are desorbed by a fall in salt concentration which in turn reduces their hydrophobic characteristics throughout the gradient. [133] HIC is mainly used as an additional technique for characterization and can resolve small changes in hydrophobicity that may not be accessible with other techniques such as, Trp oxidation, carboxy terminal micro-heterogeneity, Asp isomerization and unpaired Cys. [140] Through the mild elution conditions HIC is also well suited for preparative purification during downstream processing. [141]

Hydrophilic interaction liquid chromatography (HILIC) is a separation technique based on the use of polar stationary phases and organic-aqueous mobile phases. A minimal amount of water (around 2%) is essential to create a permanent water layer between the adsorption surface of the stationary phase and the organic part of the mobile phase (the eluent). [85] The separation mechanism of HILIC is a type of partition chromatography. Polar analytes are retained by the water enriched layer. Depending on the hydrophilic character of the analyte the elution time will thus vary. Further small increases in the water content of the eluent result in a shortening of the retention time of polar analytes. Thus, HILIC is a polar pendant to RP-LC and HIC. [85] A key barrier

in making HILIC compatible for protein separations is the dissolving solvent with a high organic content as this may lead to limited solubility and precipitation. [142] Nevertheless, HILIC has become an indispensable approach for monitoring glycans of proteins. [143] Some studies implemented the technique for oligonucleotides, separating analytes with up to 50 nt, but usage has been limited to this point. [91,144]

In ion exchange chromatography (IEC), stationary phases are used which carry electrical charges on their surface. These are, for example, anionic sulfone groups or cationic alkylated amines that are covalently or electrostatically bonded to the ion exchange resin or gel. [85] The charges on the ion exchange resin are occupied by oppositely charged ions that are mobile and can be exchanged for other ions. Ionic sample molecules must displace the counter ions occupying the ion exchange sites of the matrix in order to become bound themselves. The dissociation of the analyte can take place through increasing the ionic strength of the mobile phase and thus by preferring the concurrent exchange reaction. [135] Elution can also take place by a pH gradient, which alters the charge of the analyte and hence reducing its overall charge and then also its degree of binding to the stationary phase. For biopharmaceuticals, both anion-exchange (AEX) and cation-exchange (CEX) modes are widely employed, with CEX being typically used for proteins and AEX for NAs. [133] Samples are injected onto the column at a pH and salt conditions where the analyte's net charge is opposite the charge of the stationary phase. Given that mAbs predominantly have a basic pI and are the most important class of biopharmaceuticals, CEX is the gold standard for monitoring PTMs that cause charge heterogeneity like deamidation or aspartate isomerization. [92] The technique can be used in moderate conditions, e.g. regarding mobile phase, column temperature, and is generally robust. It is thus utilized as a validated method where it is detecting charge variants as low as 0.5% accurately e.g. for release testing of trastuzumab. [97] Due to its mild elution conditions, high binding capacity, and moderate to high chromatographic resolution, it is an efficient technique for purifying structurally and functionally intact proteins with high yields. CEX, like other adsorption methods, does not limit the amount of applied sample *per se*, alleviating the need for previous concentration stages. Nonetheless, CEX has the benefit of allowing concentrated protein fractions to be produced from dilute materials and is thus used in the purification process of mAbs. [141]

AEX can be used for separating NAs based on their negatively charged phosphate backbones. The NAs are thereby separated based on the number of their negative charges and thus their length. As an analytical technique it is limited to oligonucleotides since the differences in charge between very large NAs is too small in most cases. [91] However it has proven to be very effective for separation of NAs from proteins during downstream processing, either to remove host cell NA

from protein biopharmaceuticals [141], proteins from in vitro transcribed mRNA [145] or digested mRNA from DNA during plasmid purification [29]. Furthermore, AEX is used for purification of recombinant viruses for gene therapy products such as rAAV or lentiviral vectors. [146]

1.3.7 Electrophoretic techniques

Electrophoresis is one of the most extensively used techniques in molecular biology. The wide application as slab gel electrophoresis is unquestionably attributable to its comparatively high resolution in combination with user-friendliness, and minimal apparatus needs. [147] The fundamental concepts and processes of gel electrophoresis have not changed significantly since their establishment. The samples are placed onto a gel matrix and subsequently separated by the simultaneous effects of an electrical field, buffer ions, and the sieve-like gel. [85] When the electrophoresis has been completed, the separated molecules inside the gel are visualized with dyes, allowing qualitative or quantitative assessment. Both proteins and NAs are charged molecules that move in response to electric fields. Two essential physical characteristics for gel electrophoresis are electrophoretic mobilities and isoelectric points (pI). Analytes are influenced by their environments in terms of both of these properties. [147] The sieving features of gels have the greatest effect on migration, while pH, ionic strength and denaturants in the system are other variables that impact migration. Despite the fact that conventional types of electrophoresis are capable of resolving the elements of complicated samples, they suffer from a number of limitations. The analysis time, that is essentially limited by the amount of Joule heating, is eventually the most glaring disadvantage. Slab gels are restricted to low potential electric fields because they have a limited capacity for the dissipation of Joule heat. Furthermore, the entire procedure consists of a series of tedious, time-consuming tasks, beginning with the polymerization of the gel continuing through sample preparation and loading, separation in the electric field, and concluding with gel staining and the collection of results. Additional issues include moderate repeatability, especially for two-dimensional separations, variances in staining efficiency of different analytes, that make quantitative accuracy challenging to attain, and the absence of automation potential. [148]

Utilizing capillaries as an electromigration channel for analysis of a broad range of molecules has various benefits over the traditional solid supports. Especially capillaries with a small diameter (most frequently made of fused-silica; fs) are well suited for the purpose of electrophoresis due to their physical features. These capillaries usually have an internal diameter (I.D.) of 10–250 µm, lengths of 20–100 cm, and are externally covered with a polyimide polymer which confers great flexibility to an otherwise extremely brittle fs capillary. [85] Because of their high surface-to-

volume ratio, capillaries having these characteristics are capable of dissipating the Joule heat that is produced by strong electric fields in a very effective manner. Whereas a slab gel has a surface to volume ratio of 1.3, a conventional 50 cm capillary has a ratio of 80. [148] Due to this capacity of effectively dispersing heat, electrophoretic separations may be carried out at voltages of up to 300,000 V [149] when the capillary exterior is temperature controlled, with commercial instruments often enabling 30,000 V. Employing capillaries for electrophoresis has several additional benefits. The narrow diameters of conventional electrophoresis capillaries give total column volumes in the microliter range, enabling the use of just low milliliter amounts of buffer. Furthermore, in accordance with a general guideline limiting analyte injection to 1% to 5% of the entire capillary volume, sample quantities put into the capillary are in the nanoliter range. [148] Consequently, only 5 µL of sample may be sufficient for multiple runs. In addition, the use of such low sample volumes is rather limited by sample stability or evaporation than sample consumption per injection.

The basic CE equipment configuration consists of two buffer vessels coupled by a high voltage source. The capillary ends are submerged in the buffer solutions. One end of the capillary is temporarily put into the sample vessel in order to inject the analytes into the capillary using a

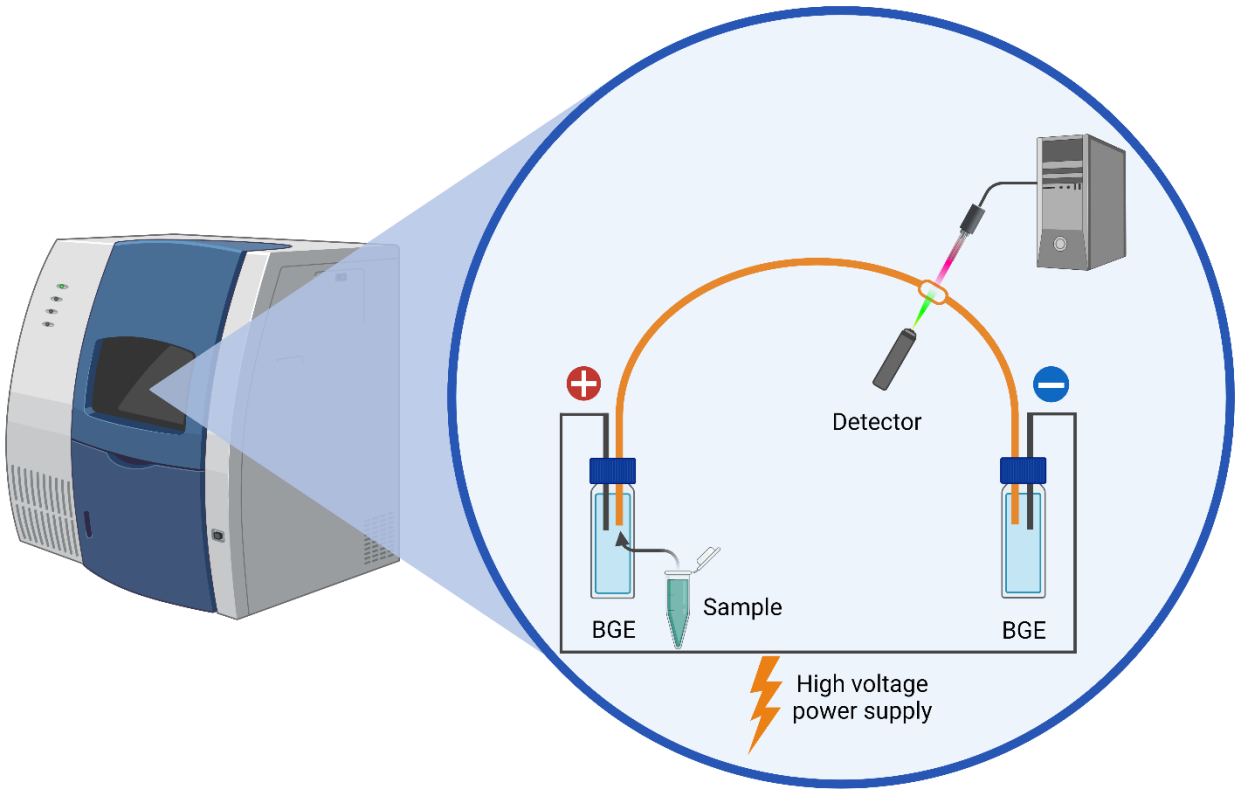


Figure 7: Schematic illustration of a capillary electrophoresis instrument with a laser induced fluorescence detector

variety of procedures (*vide infra*). In the electric field, charged molecules move to their counter electrode and pass through the detector in an electric field (see Figure 7). [148] Compared to HPLC, injection of samples into CE capillaries is more challenging since considerably lower quantities must be applied into a miniaturized environment. The following three injection procedures are favored on commercial devices. With the hydrodynamic approach, a pressure differential is established between the beginning and end of the capillary by applying pressure to the sample vessel. With increasing pressure and injection duration, more sample volume is administered. Similarly, by providing vacuum to the buffer vessel into which the capillary end is submerged, a pressure differential is generated that leads to injection of the sample. In electrokinetic injection, the capillary tip is put into the sample container. Short voltage intervals induce the injection of sample molecules. The amount injected is dependent on the voltage and time. [85] The detection of the samples may be carried out employing spectroscopic methods (UV/Vis; Fluorescence) or mass spectrometry. [147]

The electroosmotic flow (EOF) is a separation phenomenon which is highly relevant for CE, and depending on the use case it may have different effects on the separation. Capillaries made of fs feature surface silanol groups that are neutral. Above a pH value of approx. 3, hydrogen ions are released upon contact with buffers, and a negatively charged surface layer is created. [85] This causes the formation of a static double layer by attracting positively charged ions from the migration buffer. At increasing distance from this layer, cations and anions organize themselves in a less ordered format, which is why this region is known as a diffuse layer. [150] When the distance from the capillary wall increases, the affinity of positive particles for silanol groups diminishes until a layer of a balanced electrolyte is found. This exponential decline is the cause of the electroosmosis effect and is characterized by the zeta potential. Due to the cohesive nature of the hydrogen bonds between the different layers in aqueous solutions, the whole buffer solution is attracted to the cathode (see Figure 8). [148] Via this electroosmotic flow, dissolved neutral analytes and larger and smaller or singly and multiple charged anions are also pulled to the cathode. This implies that under carefully chosen conditions the analysis of all species, anionic cationic or neutral, can be conducted within a single experiment. The EOF rises as the pH of the migration buffer increases. This contrasts with most HPLC modes, where continual mass transfer of sample molecules between the mobile and stationary phases takes place. Thus under ideal conditions, the analyte`s interaction with the stationary phase is perpendicular to the flow direction which causes retardation but also contributes to peak broadening during separation. In ideal CE, a mass transfer does not occur and both driving forces, the electrophoretic movement of molecules in the electric field and the overlaying EOF, are oriented

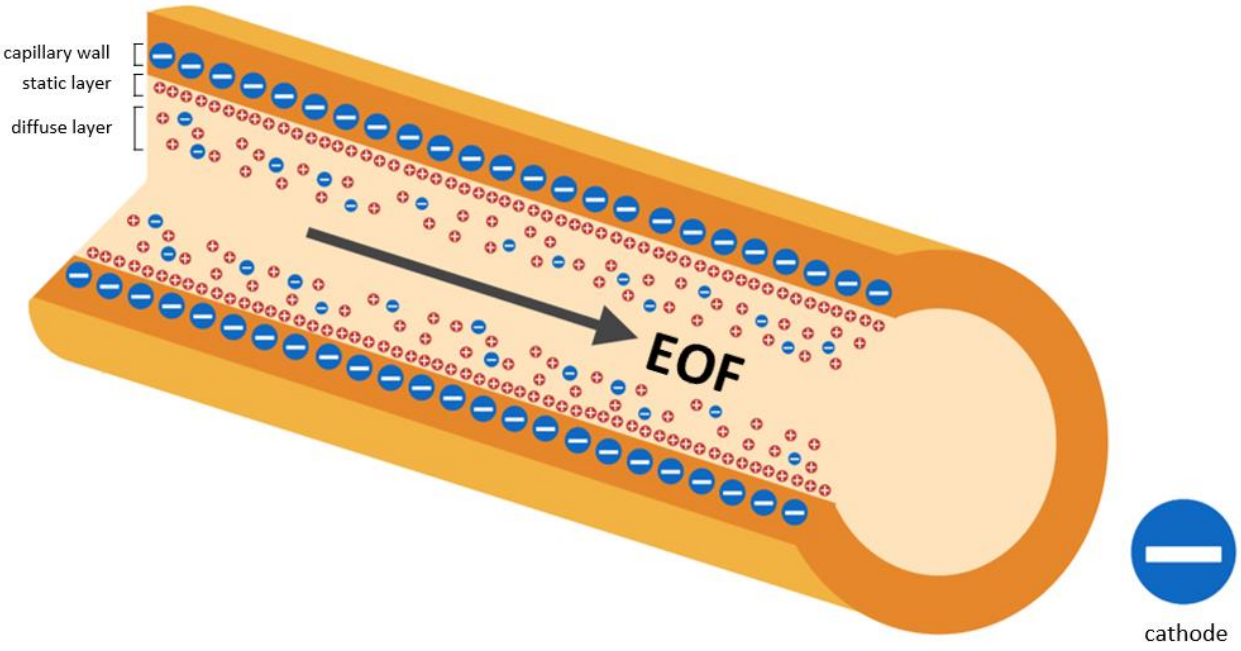


Figure 8: Cross-section of a fused silica capillary with schematic illustration of the electric layers and the electroosmotic flow

in parallel, potentially with different prefix. This results in a plug shaped profile and ultimately leads to less peak broadening and increased separation efficiency. [85] In practice, the interactions of biomacromolecules with the surface of fs capillaries, are identical to those seen on chromatographic silica substrates. Since these couplings are frequently strong and permanent, they pose significant challenges to the effective use of CE for biopharmaceutical analysis. As mentioned before, silanol groups are slightly acidic and ionize at pH higher than 3, with the net charge on the surface rising until pH 10, upon which the silanol groups are completely deprotonated. [147] In this environment, proteins with surface-positioned basic amino acid residues may engage in coulomb interactions with ionized sialons. Although fs consists mainly of silanol groups, it may incorporate hydrophobic patches on its amorphous surface. Owing to the heat treatment of the capillary during its production, siloxane is produced by the elimination of vicinal hydroxyl groups and the subsequent development of a Si-O-Si linkage. Through their minimal protonation, these structural elements have a limited affinity for water and constitute hydrophobic sites inside the fs capillary. [151] For this reason, non-polar sites on a macromolecule might also interact with the silica surface. Although the backbone of NAs is negatively charged due to phosphate groups, the neutral bases may overcome this repulsion and also adhere to the fs surface. This effect mainly affects ssNA while, dsNA adhesion is driven in large part by the creation of intermolecular hydrogen bonds in the phosphate-silica contact layer. [152,153] This

mass transfer causes peak broadening, tailing, and, in severe cases, a diminished detector response or the entire disappearance of signals. Moreover, adsorption modifies the condition of the capillary wall during an analysis and may vary the amount of EOF from one analysis to the next, resulting in poor repeatability. [151] Therefore, four different techniques frequently utilized individually or in combination to limit wall interactions have to be mentioned: Operating at high or low pH, the use of buffer additives, coating of the fs inner surface, and high salt concentrations. [147]

The family of CE techniques offers a wide variety of separation modes. It is possible to separate molecules based on charge, hydrodynamic volume, molecular weight, pI, or hydrophobicity. The very same CE apparatus is suitable for zone electrophoresis, affinity separation, gel electrophoresis, isoelectric focusing (IEF), isotachophoresis, and chromatographic procedures like micellar electrokinetic chromatography and capillary electrokinetic chromatography. The modes encountered primarily in the biopharmaceutical industry are capillary zone electrophoresis (CZE), capillary IEF (cIEF), and capillary gel electrophoresis (CGE).

In a CZE analysis, the simplest mode, the capillary is filled with an adequate separation buffer at a chosen pH, and subsequently the sample is injected. High voltage is delivered to the system while the ends of the capillary and the electrodes are positioned in buffer vials. The loaded molecules in the injection plug travel with an electrophoretic mobility (μ) which is determined by their charge to hydrodynamic radius ratio (q/r), and finally pass a detector where data is gathered. [85] By a commonly used definition, the "normal" polarity for CZE is inlet = anode, detector, and outlet = cathode, while “reverse” polarity refers to the opposite. With proceeding electrophoresis, the ionic species segregate based on their respective electrophoretic mobilities and traverse the detector as analyte zones. Thus the, the name capillary zone electrophoresis was coined. [148] The composition of the electrophoretic buffer, also referred to as background electrolyte (BGE), is the most influential chemical variable on CZE. This includes the buffering agent type and concentration, chemical modifiers and the pH value. Physical parameters comprise the capillary dimensions, temperature and the electric field strength. The pH of the BGE determines the net charges of the sample components and, consequently, their mobilities. The ideal BGE ought to have adequate buffering capacity in the pH range that offers optimum resolution, introduces low electro dispersion impact, provides a low background in the detector system, encompasses low conductivity at concentrations ensure effective pH buffering, and prevent wall interactions. [147,154] Due to the fact that sample properties determine the appropriate pH range, no single BGE can be exploited for all analytical tasks. At pH levels that are closer to the analyte's pI, reduced mobility will cause extended run durations, peak broadening,

and an increased likelihood of wall adhesion. Therefore, pH is often recommended to be at least one unit higher or lower than the pI of the target molecule. [148]

Choosing the I.D. of the capillary is a trade-off among resolution, sensitivity, and loading volume. Lowering the diameter of the capillary to improve heat dissipation yields the highest resolution, while large I.D. provide the best sensitivity and sample loading capacity. The recommended capillary I.D. for most purposes is 50 µm, although 75 to 100 µm may be necessary for high-sensitivity or micropreparative operations. [148] The length of the capillary may be increased to enhance resolution, but often has minimal impact in comparison to BGE composition. In case the BGE is not able to sufficiently minimize wall interactions, capillaries can be coated by covalent reaction of various agents to the silanol groups. Most often neutral polymers like poly vinyl alcohol or polyacrylamide are used for biological analytes. [148]

Simple CZE has limited use for NA, since q/r is very similar for NA of varying lengths. However, it was reported that oligonucleotides may be separated based on their hydrophobicity and thus their sequence. [155] Despite this, the main application of CZE is monitoring of charge heterogeneity of proteins. There are “generic” methods for mAbs [156,157] but also other prominent examples like insulin [158]. It serves as an orthogonal method to CEX and was shown to be stability indicating, precise, accurate, robust, linear and sensitive for mAbs. [156] In addition, it represents an alternative to HILIC for glycan analysis. CZE has various benefits over other techniques. It can be coupled to MS allowing peak characterization. [159,160] Sample preparation is often simple, and analysis durations are quick, allowing for high-throughput screening. [148] It is not constrained by the size of the analyte and may be used to analyze anything ranging from small molecules, mAbs, viruses, and even cells. [161] On the other side, the design of a BGE which is tailored to the analyte may be tedious in the beginning and inhibit the application. Additionally, sensitivity can be an issue, especially with UV detection and narrow capillaries. [85]

Affinity CE (ACE) is a synergistic method that combines CZE separation and affinity interaction with ligands, hence increasing the selectivity for desired charged species. By adding a ligand to the separation system, a complex with different migration characteristics than the target itself is created. This approach may be used to analyze binding reaction parameters (such as concentration, stoichiometry, and equilibrium) and complex formation processes. [162] An ACE experiment's design is highly dependent on the analytes and their binding constants (K_D). For tight binding events (K_D in the nanomolar range), the protein analyte and its antigen can be pre-incubated prior to CZE separation (equilibrium-mixture analysis). Throughout the analysis, the generated inert complex will be stable, permitting normal CZE conditions. [148] Less stable

complexes (K_D in the micromolar range) will lead, as a result of continuing dissociation throughout the analysis caused by continuously changing equilibria, to extensive zones in which the equilibrium between ligand and protein is permanently altered. The addition of the ligand to the BGE, which allows for a stable equilibrium within the capillary through a consistent ligand concentration, is an efficient strategy for such unstable complexes (mobility-shift analysis). [162] Nevertheless, not all ligands may be abundant enough to be incorporated into the BGE or might by incompatible with the separation, e.g. due its own properties or impact on the BGE properties or homogeneity. In such instances, analyte and ligand may be injected individually into the capillary (two plug approach). This needs the protein and ligand to have distinct electrophoretic mobilities so that one can catch up the other. This approach is also known as electrophoretically mediated microanalysis or flow-through partial-filling (FTPF) ACE. [163] The capillary as a microreactor offers some advantages: It reduces the consumption of precious analytes and expands the automation possibilities of the analytical method.

All modes of ACE have a wide range of application like binding interaction characterization, immunoassays, aptamer evolution and drug discovery. [164] Through its specific action mode it is mostly employed in the research and characterization phases of drug development, but it was also described for charge heterogeneity profiling of mAb mixtures [165] or coupled with MS to monitor oxidation and binding affinity of mAbs simultaneously. [166]

Capillary isoelectric focusing (cIEF) unites the resolution of traditional gel isoelectric focusing (IEF) with the benefits of CE. As the pI is nearly the same for all kinds of NAs, this technique is almost exclusively used for protein charge heterogeneity profiling. IEF refers to electrophoresis that is carried out in a pH gradient. When an IEF separation is conducted, the amphoteric compounds in the sample are separated and concentrated at spots along the pH gradient where their pI is equivalent to the pH. [147] As with other CE techniques, making use of capillaries provides effective Joule heat dissipation and enables the application of high voltage, which speeds up the process of concentrating protein zones. In cIEF, the capillary is loaded entirely with a combination of carrier ampholytes (CAs) and the target molecule, which represents the test sample. A separation voltage is provided to the two ends of the capillary, which are dipped into the catholyte and anolyte, respectively. Through isoelectric stacking of components in the CAs, a pH gradient is formed from the anodic to the cathodic side of the capillary. Proteins in the mixture travel to sites within the capillary where their pIs are analogous to the pH in the pH gradient. As they are uncharged when reaching the pH that corresponds to their pI, their migration ceases. This in turn creates compact focused protein zones, which can then be detected. [85] There are two fundamental prerequisites for an efficient cIEF separation column: no flow (electroosmotic

or hydrodynamic) and no capillary wall adhesion. Therefore, capillaries are typically coated with stable, neutral coatings to suppress EOF and avoid protein contact with the silica surface. [148] To avoid hydrodynamic movement during the focusing, the two tips of the capillary should be kept on a single level. When the two prerequisites are met, a stationary gradient can be established where the protein zones are focused in narrow zones. Due to the fact that generic CE devices are typically equipped with a single point detector the static zones inside the capillary must be mobilized in order to pass the detector, adding an additional step after focusing. Three strategies to mobilize are available, which might be combined in any way: salt mobilization, pressure-driven mobilization, and EOF-driven mobilization. [147] The pressure-mobilization procedure might lower resolution and increases sample analysis times. The chemical mobilization process, on the other side, requires more method development, but may increase resolution through additional separation effects. Nowadays, the industry standard is whole-capillary detection, in which the IEF over the whole length of the separation column is observed by a UV/Vis detection system usually operated at 280 nm, to avoid CA background noise. [167] These instruments are commonly referred to as imaged cIEF (icIEF). Nevertheless, mobilization may still be necessary for fractionation or MS-coupling.

Just like BGE in CZE, the CA is the most important condition in cIEF. Given that a CA may comprise more than 900 amphoteric compounds, it is mainly obtained commercially. Owing to the vast number of CA ingredients, background absorption is a special concern aside from factors like optimal resolution. Therefore, a compromise between buffering capacity of the ampholyte and UV absorption must be made. [147] After focusing proteins are concentrated by several magnitudes into compact zones at net-zero charge, which poses the risk of precipitation and aggregation. Therefore, sugars, surfactants, and chaotropic salts can serve as additives to the CA to preserve a protein’s solubility at their pI. Moreover, polymers like methyl cellulose stabilize the capillary surface, further reduce surface interactions, and improve resolution by decreasing protein diffusion coefficients. [147] When a sample's pI value is uncertain, a CA with a broad pH range (e.g. 3-10) is often the first option for pI screening. For most analytes, separation resolution may be improved in successive runs by using CAs with a tighter pH range. The focusing time is an additional relevant factor. Stabilization of the focusing current essentially corresponds to conclusion of CA focusing, whereas target proteins may travel more slowly than CAs, often necessitating experimental assessment of the focusing time. [148] In QC labs, the icIEF with UV detection has become a standard technique for charge heterogeneity investigation. It is a rapid approach that has been confirmed in inter-laboratory investigations for its robustness and repeatability for charge heterogeneity profiling of therapeutic antibodies. [168] It is particularly

valuable due to its practically universal application to all kinds of proteins and simple method development, which has been standardized by commercial suppliers. [169] Coupling with MS has been described but is often complicated by the need to remove or use uncommon Cas. [170] In this situation, CEX and CZE are often preferred alternatives. Additionally, some proteins may be difficult to analyze due to precipitation, even when countermeasures were taken.

The implementation of capillaries eliminated the majority of difficulties related with gels in electrophoresis, such as the removal of convection due to the quick Joule heat dissipation and the reduction of diffusion owing to the strong electric field. However, gels have the potential to actively contribute to the separation by acting as a sieving matrix that selectively impacts the mobility of analytes according to molecular size. Throughout CZE, biomolecules like NAs and SDS–protein complexes do not display any substantial mobility variances, regardless of their length, due to their constant q/r. Thus, the inclusion of an engaging sieving medium is required to achieve resolution. For the initial implementations of capillary gel electrophoresis (CGE), cross-linked high-viscosity gels polymerized within the capillary predominated. [171] Newer techniques use linear, UV-transparent polymers, such as dextrane, cellulose derivatives, or polyethylene glycol, for simpler and robust handling, improved repeatability, and more injection mode flexibility. [148] The formation of an entangled superstructure in a solution of noncross-linked polymers can be observed above a certain concentration threshold and forms the foundation of the sieving effect. Intermolecular polymer and polymer–solvent interactions are shaping the polymer network structure, which is further influenced by factors such as temperature, polymer concentration, molecular weight (M_w), shape, and additives like SDS. [172] The network structure is dynamic, with channel-like structures in the polymer continuously emerging and vanishing. Therefore, the theoretical pore size is flexible and solely reliant on the concentration of a given polymer. In general, increased polymer concentrations facilitate the resolution of analytes of lower molecular mass. [147] Moreover, the use of an electric field may assist in polymer organization. In contrast to the fixated cross-linked gels, the dynamic pore structure enables even the largest molecules to permeate. According to the Ogsten model, analytes with a size similar or lower to the pore size are resolved by a true sieving process, and their mobility is proportional to $\exp(-M_w)$. [173] Large Stokes radii molecules, such as the flexible chain of biopolymer molecules, might well transit through holes far narrower than their size would allow. This is described by the reptation model (reptation-without-stretching regime), which characterizes macromolecule movement as a "head-first, snakelike" flow through the pores of the sieve medium. According to this theory, the mobility of the analyte is inversely related to its M_w. This holds true up until the point when

maximal size is reached, at which a repetition-with-stretching process leads to a ceiling for mobility, and the analytes are not distinguishable. [148]

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is the most used slab gel electrophoresis technique for the study of polypeptides. Correspondingly, CE-SDS, its capillary equivalent, represents the most abundant member of the CE family used in the biopharmaceutical industry. SDS adheres to peptide chains in an essentially stoichiometric manner, with about one SDS molecule attached for every two amino acid residues. [172] Considering that the impacts of ionic functional groups are negligible compared to those of the detergent sulfate groups, the SDS–protein complexes exhibit a comparable q/r regardless of the length of the polypeptide chain. Incompletely complexed proteins may display varied mobility and may be resolved into numerous species, while with smaller proteins, the inherent charge of the side chains has a greater impact, resulting in larger variability despite similar M_w. The surfactant's solubilizing ability is a significant benefit of utilizing SDS, permitting the examination of proteins that are prone to precipitation. [147] While protein separations via other CE modes is often adversely impacted by capillary wall interactions, this issue is negligible or absent in CE-SDS. The BGE pH is adjusted so the fs surface has a negative charge, and the SDS–protein combination is anionic at virtually all pH levels, leading to electrostatic repulsion and consequently preventing adsorption. [148] For sample preparation, proteins are diluted in a basic-pH SDS-containing buffer and subsequently heat-denaturated. Prior to heating, a suitable reducing agent such as β-mercaptoethanol or dithiothreitol is added if the proteins are to be investigated under reduced conditions. In instances when increased sensitivity is needed, fluorophore derivatization may be advised. [147] Due to their practically constant q/r, SDS–protein complexes are injected with an identical effectiveness. In contrast to CZE, where electrokinetic injection may lead to a bias since lower mobility analytes will move more slowly into the capillary, resulting in decreased relative concentrations within the injected zone. When choosing a high ionic strength for the BGE, this can be exploited for a stacking effect, which sharpens the injected zone and increases sample concentration. [147] Electrokinetic injection is hence a favorable option, but if the sample has a significant salt content and desalting is impractical, hydrodynamic injection may be utilized. Another difference to CZE is the high viscosity of the run buffer which necessitates adequate pumping time to fill up the capillary.

CE-SDS is a recognized technique for the examination of protein biopharmaceuticals, is included in a large number of regulatory filings and considered a “generic” method which matches many different analytes. [172] While it is conceivable to build a CE-SDS method by testing multiple gel buffer compositions, this is often considered a substantial investment, and in practice the

majority of laboratories use commercially available analysis buffers. Via the non-reduced form of the technique fragments, covalently bound aggregates and host cell proteins are assessed, while with the reduced mode serves for monitoring of fragments and glycan distribution. [172] Due to its extensive usage, high throughput applications of CE-SDS, such as in short capillaries with minimal analysis times or with multichannel devices, are also commercially available. [174] As SDS interacts with ESI, the principal downside of this method is the limited ability to characterize peaks using MS. The removal of SDS by a two-dimensional CGE-CZE-MS method has been established but remains a technically demanding niche application with great potential. [175]

CGE was established as a method for DNA analysis and sequencing in the early 1990s and continues to be one of the leading methods for the purity assessment of NAs. The separation technique is identical to CE-SDS, and comparable gel matrices may be used. It is applicable to the whole range of NAs, from single bases to tens of thousands of bases. [173] A difference is the frequent use of intercalating dyes such as ethidium bromide which raise the resolution and sensitivity through fluorescence detection by many orders of magnitude. [176] The approach is utilized orthogonally to IP-RP-HPLC for the sizing of oligonucleotides, but is particularly beneficial for longer NAs like mRNA as very few techniques can resolve them. In recent years, CGE has regained attention in the separation of NAs for detection of contaminants, identification of modifications, or the investigation of secondary structures as a result of the biopharma industry's shift from mAbs to other modalities. It has been successfully applied for gene therapy products and mRNA vaccines. [91] Due to BGE components that are incompatible with MS, CGE-MS applications for NAs have been scarce so far.

2. Results and Discussions

The objective of this dissertation is to explore the potential of capillary electrophoresis as an analytical technique to evaluate the properties of novel therapeutic agents, including peptides, plasmids, mRNA/LNP formulations, and rAAV. The ultimate intent is to expand the applicability of CE to all these modalities as much as possible. The doctoral thesis encompasses four focus areas that were addressed in a total of five articles that were published in peer-reviewed analytical chemistry journals. Each of the sections of the thesis will be presented through a summary of each project, along with the reprint of the corresponding articles.

In the initial section (2.1), a technique is presented for identifying and measuring the degree of methionine oxidation in proteins, employing Ag(I) and Au(III) ions as affinity probes. The process of methionine oxidation is a frequently occurring PTM that has the potential to impact the structure and function of proteins. [21] The article presents an investigation on the capability of ACE to quantify the degree and position of methionine oxidation in peptides and offers complementary approach to MS.

Section two (2.2) provides a direct comparison of the present analytical methodologies employed for the analysis of rAAVs, a highly promising vector for gene therapy. The rAAV is a type of small, non-enveloped ssDNA virus that has the ability to deliver therapeutic genes into target cells. [177] It exhibits numerous benefits in comparison to alternative gene transfer vehicles but presents various analytical characterization challenges which are owed to the combinations of viral proteins and NAs. The research presented in this chapter delineates the benefits and challenges associated with employing CGE, CZE, TEM, AUC, and IP-RP-LC methodologies for the purpose of rAAV characterization. The findings demonstrated that CE techniques are capable of providing distinctive insights into rAAV, which may not be readily accessible through other analytical approaches.

Chapter three (2.3) presents an approach for quantification and identification of NA therapeutics through the utilization of fluorescent peptide nucleic acid (PNA) hybridization probes for CE. PNA probes are artificially synthesized, neutral polymers that have the ability to selectively hybridize with target NAs [178] and induce changes in their electrophoretic mobility. The article demonstrates the application of PNA probes in the qualitative and quantitative analysis of NA therapeutics, including oligonucleotides, rAAVs, and mRNA.

The final chapter (2.4) describes a study which demonstrates different analytical methods for longer NAs (> 500 nt), highlighting their specific advantages and limitations. This involves, the PNA based ACE method (chapter 2.3), ddPCR, CGE (chapter 2.2) which was expanded by the use of nuclease digestion, IP-RP-LC and CZE. These methods were applied to a variety of NA formulations: mRNA, rAAV and LNPs.

2.1 Methionine oxidation of proteins analyzed by affinity capillary electrophoresis in presence of silver (I) and gold (III) ions

Methionine (Met) is an amino acid that shows susceptibility to oxidation by reactive oxygen species, leading to the formation of methionine sulfoxide (MetO). [179] The level of Met oxidation in a protein is regarded a CQA, due to its potential to cause reduced potency, aggregation and altered pharmacokinetics depending on the position in the protein´s amino acid sequence. [180] Peptide mapping via LC-MS is considered the gold standard for addressing this CQA. [181] However, this approach requires costly equipment, can be complex and is a time-intensive process. For this reason, the adoption of ACE as a technique for separating peptides and proteins with varying oxidation states is proposed in this study. This is achieved through the interaction of biomolecules with ligands present in the BGE. This methodology's feasibility was shown through peptides with differing Met oxidation variations as model analytes. The metallic cations of Ni, Pd, Pt, Cu, Ag, and Au were subjected to testing, and the utilization of Au and Ag yielded the most favorable outcomes. The ions of Ag(I) and Au(I) have the property of exhibiting selective complex formation with thioethers as opposed to sulfoxides. The introduction of the mentioned ions into the BGE results in a selective formation of complexes with Met residues. This leads to a modification in the charge, enabling the separation of species based on their individual oxidation state of Met. Experimental data indicated that gold in its trivalent state (Au(III)) undergoes reduction to its monovalent state (Au(I)), while Met oxidizes to MetO within the capillary. Au(I) has the ability to form complexes with peptides that contain Met and can thus alter their mobility in ACE. For this reason, the application of $AuCl_3$ in conjunction with thioether containing samples is suitable for investigating the conduct of Au(I) in CZE at concentrations approximately ranging from 250 to 500 µM. Furthermore, $AuCl_3$ may be a noteworthy oxidizing agent for inducing stress in biological specimens. The specific binding of Ag(I) with thioethers has been effectively utilized in ACE to differentiate between peptides having varying levels of oxidation in their Met residues. The investigation has demonstrated a distinct differentiation between non-oxidized, singly-oxidized, and doubly-oxidized variants, which were previously indistinguishable through conventional CZE. The data obtained indicated that there is a binding interaction between a single Ag(I) ion and a single Met residue, leading to an increase in the net charge of the analyte. The logarithmic binding constant (log K_B) of 8-mer peptides that contain a single Met residue was measured to be approximately 3.6, which is consistent with literature. Studies involving mAbs have demonstrated that the interaction between Ag ions and large proteins can impact their mobility and that the method possess the capability to serve as an indicator for protein oxidative stress.

1ST PROJECT:

Methionine oxidation of proteins analyzed by affinity capillary electrophoresis in presence of silver (I) and gold (III) ions

Electrophoresis. 2021 Jun; 42(11):1209-16.

Andrei Hutanu[1,2]
Peter C. Hauser[2]
Bernd Moritz[1]
Steffen Kiessig[1]
Aurélie Noël[1]
Jan O. Stracke[1]
Markus Wild[1]
Maria A. Schwarz[2,3]

[1]Pharma Technical Development Europe (Biologics) Analytics, Basel, Switzerland
[2]Department of Chemistry, University of Basel, Basel, Switzerland
[3]Business Unit Biopharmaceuticals, Solvias AG, Kaiseraugst, Switzerland

Received December 2, 2020
Revised January 25, 2021
Accepted February 12, 2021

Research Article

Methionine oxidation of proteins analyzed by affinity capillary electrophoresis in presence of silver(I) and gold(III) ions

Oxidative damage of biopharmaceuticals during manufacturing and storage is a key concern throughout pharmaceutical development. However, few simple and robust analytical methods are available for the determination of oxidation sites. Here, the potential of affinity capillary electrophoresis (ACE) in the separation of proteins with oxidized methionine (Met) residues is shown. Silver(I) and gold(I) ions have the attribute to selectively form complexes with thioethers over sulfoxides. The addition of these ions to the BGE leads to a selective complexation of Met residues and, thus, to a change of charge allowing separation of species according to the different oxidation states of Met. The mechanisms of these interactions are discussed and binding constants for peptides containing Met with silver(I) are calculated. Additionally, the proposed method can be used as an indicator of oxidative stress in large proteins. The presented technique is easily accessible, economical, and has rapid analysis times, adding new approaches to the analytical toolbox of Met sulfoxide detection.

Keywords:
Affinity capillary electrophoresis / Gold / Methionine sulfoxide / Monoclonal antibody / Silver
DOI 10.1002/elps.202000355

Additional supporting information may be found online in the Supporting Information section at the end of the article.

1 Introduction

Damage to proteins caused by oxidation is a commonly observed problem in biopharmaceutical development, mainly affecting methionine (Met), cysteine (Cys), histidine (His), and tryptophan [1, 2]. Unlike the side chains of the branched chain amino acids, the Met side chain is rather flexible but possess a similar hydrophobicity [3, 4]. This makes Met unique among the proteinogenic amino acids, allowing it to adapt its shape to nonpolar structures. Met oxidation caused by reactive oxygen species leads to the formation of the polar methionine sulfoxide (MetO). *In vivo*, this modification is altering protein structures and accelerating age- dependent diseases [5, 6]. For proteins stored *in vitro*, oxidation of Met residues to MetO causes loss of potency and aggregation [7–10]. Monoclonal antibodies (mAbs), which are of special interest to the pharmaceutical industry, tend to have altered pharmacokinetic properties, due to Met oxidation [11]. The effects caused by MetO are regarded as critical quality attributes. Hence, special techniques for the detection of this posttranslational modification are required.

As an antioxidant defense, cells use methionine sulfoxide reductases, which catalyze the reduction of MetO using thioredoxin. For the monitoring of MetO *in vivo*, recombinant methionine sulfoxide reductases labeled with a fluorescent marker can be used [12]. For chromatographic detection, the chemical attributes of MetO can be exploited: It is more polar, bulkier, and less flexible than Met [13–15]. This change of attributes may lead to a small increase in polarity if Met is located on the exterior of the protein. However, the polarity difference is typically too small for a direct separation with HPLC. Only if the oxidation takes place at a critical Met residue and, thus, leads to a conformational change, does the resulting species have a completely changed hydrophobicity. Oxidation variants can then be separated in RP-HPLC or

Correspondence: Andrei Hutanu, Pharma Technical Development Europe (Biologics) Analytics, F. Hoffmann-La Roche AG, Grenzacherstrasse 124, 4070 Basel, Switzerland.
E-mail: andrei.hutanu@roche.com
https://orcid.org/0000-0003-0174-5250

Abbreviations: **2M**, peptide with sequence: YAMAAMKA; **CQA**, critical quality attribute; **Cys**, cysteine; **DAPS**, *N*-Dodecyl-*N,N*-dimethyl-3-ammonio-1-propanesulfonate; **EACA**, ϵ-aminocaproic acid; **His**, histidine; **IGF-I**, insulin-like growth factor I; **mAb**, monoclonal antibody; **Met**, methionine; **MetO**, methionine sulfoxide; **M-(3)-*O***, peptide with sequence: YAM(ox)AAMKA; **M-(6)-*O***, peptide with sequence: YAMAAM(ox)KA; **MOMO**, peptide with sequence: YAM(ox)AAM(ox)KA; **MXO**, M-(3)-*O* and M-(6)-*O*; **TETA**, triethylenetetramine

hydrophobic interaction chromatography [16, 17]. In the case of a reduced affinity of a protein to its target caused by critical residue Met oxidation, an affinity chromatography can be used to separate a protein [11]. Nevertheless, these approaches can just work for certain classes of proteins and are not applicable in a generic way. For an overview of all Met residues typically a proteolytic digest is performed. In the resulting peptides, the increased polarity of an oxidized residue is of a greater significance than in a protein, allowing separation with RP-HPLC. Coupling of LC with MS was already introduced in the early 90's and helps to identify all oxidized Met in a protein [18]. With the development of high throughput proteomics, the identification of MetO is now possible for whole cell extracts [19]. However, this procedure is relatively time consuming and complex.

Changes in polarity by Met oxidation do not typically have a great impact on the electrophoretic mobility in CE. However, Nashabeh and colleagues [20] successfully separated Met oxidation variants of the 70 amino acid long insulin-like growth factor I (IGF-I). In a polyacrylamide-coated capillary with an acidic buffer composed of β-alanine, citric acid, zwitterionic *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate (DAPS) and acetonitrile, hydrophobic interactions between DAPS and the protein separated the Met and MetO variants of IGF-I. In this study, acetonitrile inhibited micelle formation in the BGE leading to a RP-HPLC-like separation mode.

The addition of metal ions to the BGE has been shown in a wide range of studies to be an appropriate method for studying interactions of biomolecules including DNA and proteins [21–24]. Here, we have performed a feasibility study on identification and quantification of Met oxidation variants by using ACE with metal ions as complexing reagent for nonoxidized Met. In our study, we choose to focus on silver and gold salts. As soft acids, silver and gold tend to strongly bind to soft bases such as unreduced sulfur. This interaction is known to be several orders of magnitude higher than to nitrogen and oxygen containing compounds [25]. Since the 1940s, thiol concentrations in amino acid and protein containing solutions were determined via amperometric titration in the presence of silver compounds [26–28]. Gold interactions with Met were mostly studied in the context of developing new (anticancer) drugs [29]. To best of our knowledge, the usage of both ions for the detection and measurement of Met oxidation by ACE has not yet been reported. We describe the reaction mechanisms between peptides containing Met and Ag/Au ions that are proposed to take place during electrophoretic separation. Binding constants between these peptides and Ag are calculated and the utility of the method for an estimation of protein oxidation is demonstrated.

2 Theory of ACE

ACE is a highly efficient tool for binding studies. For the complex formation [PM] between a peptide analyte P with a metal ion M, a simple 1:1 equilibrium (Eq. 1) and the corresponding mass action law can be formulated (Eq. 2):

$$P + M \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} PM, \tag{1}$$

$$K_B = \frac{1}{K_D} = \frac{k_{on}}{k_{off}} = \frac{[PM]}{[P]\,[M]}, \tag{2}$$

where k_{on}, k_{off}: rate interaction constants of the complex formation and dissociation, respectively. K_B: binding constant of the complex; K_D: dissociation constant of the complex; and **[PM], [P], [M]**: equilibrium concentrations of PM, P, and M.

Most systems with weak to moderate interactions show fast kinetic reactions. In CZE, a stable equilibrium between analyte and ligand can be reached by adding the ligand to the BGE. If the ligand changes the charge of the analyte, as can be expected for a metal ion, the analyte migration time will be shifted according to the concentration of the ligand in BGE. The proportion of the complex in the total analyte population is given by its mole fraction χ_{PM} (Eq. 3):

$$\chi_{PM} = \frac{[PM]}{[P] + [PM]}. \tag{3}$$

The observed mobility of the analyte μ is a result of the mobility of the complex μ_{PM} and the mobility of the free analyte μ_P and their mole fraction in equilibrium [30] (Eq. 4):

$$\mu = (1 - \chi_{PM})\ \mu_P + \chi_{PM}\mu_{PM}. \tag{4}$$

Insertion of Eqs. (2) and (3) in Eq. (4) yields Eq. (5):

$$\mu = \frac{\mu_P + K_B\,[M]\,\mu_{PM}}{1 + K_B\,[M]}. \tag{5}$$

Since μ_P and $[M]$ are known, while μ_{PM} can be determined by the saturation plateau of the curve, the resulting binding curve can be used to estimate K_B via nonlinear regression. For a more sophisticated stepwise 2:1 metal to peptide stoichiometry, the following equilibrium can be formulated (Eq. 6):

$$P \underset{-M}{\overset{+M}{\rightleftarrows}} PM \underset{-M}{\overset{+M}{\rightleftarrows}} PM_2. \tag{6}$$

In this case, Eq. (5) is extended to Eq. (7):

$$\mu = \frac{\mu_P + K_{B1}\,[M]\,\mu_{PM} + K_{B1}K_{B2}[M]^2\mu_{PM_2}}{1 + K_{B1}\,[M] + K_{B1}K_{B2}[M]^2}, \tag{7}$$

where μ_{PM_2}: mobility of the double bound complex; K_{B1}: binding constant of the single bound complex; and K_{B2}: binding constant of the double bound complex.

In order to estimate the electrophoretic mobility from raw data, one can use Eq. (8):

$$\mu = \frac{L_t L_d}{U}\left(\frac{1}{t_A} - \frac{1}{t_{EOF}}\right), \tag{8}$$

where L_t: total capillary length; L_d: length to detector; U: System voltage; t_A, t_{EOF}: migration time of analyte and EOF, respectively.

3 Materials and methods

3.1 Reagents

NaOH (Cat. no. 28 244.295) and orthophosphoric acid (Cat. no. 20 626.292) were from VWR Chemicals (Radnor,USA). HCl (Cat. no. 387 800 010) was purchased from Acros organics (Geel, Belgium) and thiourea/HCl cleaning solution (Cat. no. 51 350 102) from Mettler Toledo (Columbus, USA). The following chemicals were from Sigma-Aldrich/Merck KGaA (Darmstadt, Germany): AgF (Cat. no. 226 858), $AuCl_3$ (Cat. no. 379 948), catalase from bovine liver (Cat. no. C9322), $CuCl_2$ (Cat. no. 203 149), ϵ-aminocaproic acid (EACA) (Cat. no. A2504), hydrogen peroxide (Cat. no. 16 911), histidine (Cat. no. 53 319), hydroxypropyl methylcellulose (HPMC) (Cat. no. H7509), $NiCl_2$ (Cat. no. 451 193), $PdCl_2$ (Cat. no. 205 885), $PtCl_4$ (Cat. no. 79 840), sodium phosphate dibasic (Cat. no. S0876), and triethylenetetramine (TETA) (Cat. no. 90 460). Water of HPLC grade was prepared in a Milli-Q-Station (Merck Millipore/Merck KGaA). Solutions were filtered through 0.2 μm membrane filters (Corning, New York, USA).

Peptide variants with the base sequence YAMAAMKA were synthesized by JPT Peptide Technologies (Berlin, Germany). This nonoxidized base sequence is referred to as 2 M. The variant YAM(ox)AAM(ox)KA is abbreviated as MOMO. Single oxidized species YAM(ox)AAMKA and YAMAAM(ox)KA are called M-(3)-*O* (Met3 oxidized) and M-(6)-*O* (Met6 oxidized), respectively. When referring to M-(3)-*O* and M-(6)-*O*, the abbreviation MXO is used. The C- and N-terminus are unmodified. Isoelectric point of all peptide species: 8.5. mAb1 and mAb2 were obtained internally (F. Hoffmann La Roche, Basel, Switzerland). Properties of both monoclonal antibodies (mAbs): molecular weight: ≈150 kDa; isoelectric point: ≈8.5.

3.2 Sample/BGE preparation

For peptide analysis, metals were diluted in BGE (50 mM Na_2HPO_4, adjusted to pH 3 ± 0.05 with H_3PO_4 or HCl using a SevenExcellence pH/Ion meter S500 (Mettler Toledo) from 2 to 2000 μM, with a 1:2 dilution row (1 + 1 part). The samples were a mix of 1 mM peptide(s) and 0.5 mM His, used as a marker, in water. Regarding mAb experiments, AgF was diluted at 500 μM in mAb BGE (400 mM EACA, 2 mM TETA, 0.05% HPMC pH 5.7 ± 0.05). Preparation of mAb BGE is described elsewhere [31, 32]. mAb samples were stressed at 1 mg/mL with 1% H_2O_2 at 25°C. The reaction was stopped after 1 or 6 h with 20 U of catalase. The formulation of the mAbs includes His, which can be used as a marker.

3.3 Apparatus

All experiments were carried out on a SCIEX PA800 plus System (Brea, USA) which was equipped with an UV detector, a 214 nm filter (Cat. no. 144 437; SCIEX), a temperature controlled auto sampler (±2°C), and a 30 kV power supply. Fused silica capillaries from Molex (Lisle, USA) with I.D. of 50 μm, 20 cm length from the inlet to the detection window, and a total length of 30 cm were used at 20°C. Samples were stored in the autosampler at 10°C and injected at 0.5 psi for 10 s. Polarity was positive (capillary inlet) to negative (capillary outlet) with separation voltage set at +20 kV. The currents observed under the described conditions were at around 60 μA (peptide BGE) or 25 μA (mAb BGE). Instrument control, data acquisition, and data evaluation were performed with 32 Karat 10.1 software (SCIEX). Origin software (2019 version, OriginLab Corporation, Northampton, USA) was used for data evaluation.

Before each sample injection, capillaries were flushed with 0.1 M HCl for 1 min and equilibrated for 1 min with separation buffer (peptide BGE/mAb BGE). All rinsing steps were performed at 60 psi. Fresh capillaries were primed for separation buffer conditions by five runs with a peptide or mAb sample. After ACE separations, silver-coated electrodes were cleaned by putting them in the thiourea cleaning solution for 1 h. Gold oxide coating on electrodes was removed with 1 M NaOH overnight. Capillaries were used for max. 20 injections to avoid metal complex adhesion.

4 Results and discussion

The experiments conducted were based on the hypothesis that some metal ions selectively bind Met over MetO. Adding these metals to the BGE in CE analysis would, thus, only change migration of Met containing samples whereas oxidized samples would not be affected. Literature research has brought the elements of group 10 (Ni; Pd; Pt) and 11 (Cu; Ag; Au) of the periodic table to our attention in this regard [33]. Experiments were conducted using a short peptide with the sequence YAMAAMKA. This peptide was used in the four possible Met oxidation variants (2M, M-(3)-*O*, M-(6)-*O*, MOMO, see Section 3 for details). As a marker, for detection of a possible mobility shift, a molecule that would ideally be uncharged and not interact with the BGE/metal ions had to be found. However, a BGE with pH 3 leads to a very low EOF, and uncharged molecules would not be detected in the migration timeframe of the peptides. Thus, His as a charged molecule was chosen, because it maintains its mobility at a broad range of metal concentrations (up to 4 mM tested) despite increasing ionic strength in the BGE. For practical calculations with Eq. (8), the detection peak of His was used and not the migration time of the EOF. The mobility is referred to as the relative mobility (rel. μ)

4.1 Mechanism of metal complexation of methionine

For an initial screening of the effects of metal ions on mobility of peptides, 2 M and MOMO were mixed each at 1 mM

and separation in 50 mM phosphate buffer at pH 3 was carried out as a standard without observing a clear separation of the species. In a next step, the metal ions were added at concentrations of 0.25–2 mM to the BGE. Counter ions for the metals, were chosen on the basis of solubility in water. A significant alteration of mobility could be obtained for $AuCl_3$ and AgF, but not for $NiCl_2$, $PdCl_4$, $PtCl_4$, and $CuCl_2$ (not shown). Addition of AgF to the BGE leads to a clear shift of 2 M, M-(3)-*O*, and M-(6)-*O*, which appears to be stabilized at concentrations ≥250 μM (see Fig. 1A). Under these conditions, a separation of the oxidized, single oxidized and nonoxidized species can be achieved. This indicates that the binding of Ag(I) ions to 2 M, M-(3)-*O*, and M-(6)-*O*, strongly relies on the thioether group since this is the only difference between the peptides.

Ag(I) has a strong affinity for thioether groups and is reported to form a wide range of coordination geometries, most of them ranging from linear twofold coordinate to tetrahedral complexes [34]. Sigel et al. conducted studies on dimethyl sulfide and dimethyl disulfide with several metal species including Ag(I) ions [35]. In that study, the equilibrium constants of the soft Ag(I) ion were reported to be 105 times larger for thioethers than for hard ions like Ca(II). Luehrs and colleagues found, that dimethyl sulfoxide does not form any complexes with Ag(I) in aqueous solutions, as hydrogen bonding is dominant [36].

Experiments with Met have shown that silver forms [Ag-Met] complexes below pH 4.5, that solely coordinate through the thioether group [37]. At higher pH, a [Ag-$(Met)_2$] complex is detected that coordinates also through the amino group, which was also confirmed in Met-containing peptides [38]. This suggests the binding of two Ag^+ ions to 2 M and one Ag^+ ion to M-(3)-*O* and M-(6)-*O* at pH 3 according to Eqs. (9) and (10).

$$MXO + Ag^+ \underset{K_D}{\overset{K_B}{\rightleftarrows}} \left(Ag\,(MXO)\right)^+, \tag{9}$$

$$2M \overset{+Ag^+}{\underset{K_{B1}}{\rightleftarrows}} \left(Ag\,(2M)\right)^+ \overset{+Ag^+}{\underset{K_{B2}}{\rightleftarrows}} \left(Ag_2\,(2M)\right)^{2+}. \tag{10}$$

To prove this theory, we conducted mobility shift studies using AgF concentrations from 2 μM to 2000 μM in order to estimate the binding constants in the capillary. For single oxidized peptides, Eq. (5) can be reformulated (Eq. 11):

$$\mu = \frac{\mu_{MXO} + K_B\left[Ag^+\right]\mu_{[Ag(MXO)]}}{1 + K_B\left[Ag^+\right]}. \tag{11}$$

Fitting mobility plots with Eq. (11) results in accurate fits for M-(3)-*O* and M-(6)-*O*, thus, confirming the [Ag(MXO)] complex (Fig. 2). Formation constants for [Ag(MXO)] are approximately log $K_B \approx 3.6$ (Fig. 2, Inset table). Measurements conducted with potentiometric titration on pure Met and Met-containing dipeptides, have found a very similar value of log $K_B \approx 3.2$ [37, 39]. Assuming a second-order equilibrium for 2 M, we have estimated that K_{B1} is an average of K_B(M-(3)-*O*) and K_B(M-(6)-*O*) and that the single bound species [Ag(2M)] has the same migration time as the average of M-(3)-*O* and M-(6)-*O*. Eq. (7) then results in Eq. (12):

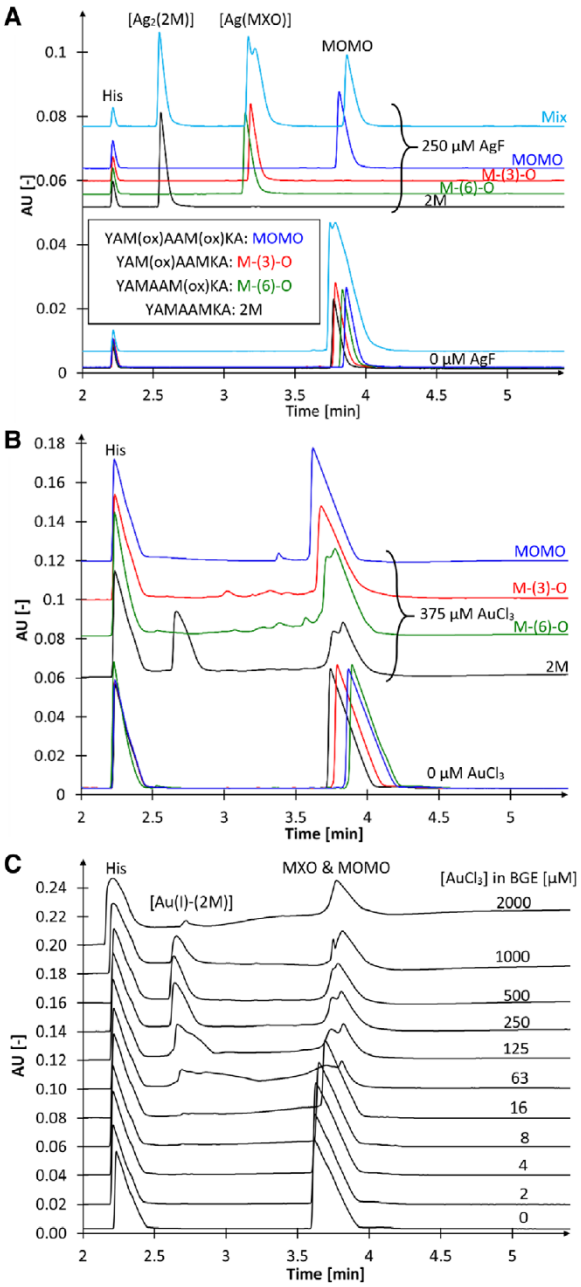


Figure 1. Capillary electropherograms of the peptides 2 M (black lines), M-(3)-*O* (red lines), M-(6)-*O* (green lines), MOMO (blue lines), or a mix of all four (teal lines) in 50 mM phosphate buffer, pH 3. (A) Addition of 250 μM AgF into the BGE. (B) Addition of 375 μM $AuCl_3$ into the BGE. (C) Effects on the electropherogram of 2 M with 0–2000 μM $AuCl_3$ in the BGE. All lines were shifted according to the marker peak (His). Separation voltage of 20 kV; capillary: fused silica.

$$\mu = \frac{\mu_{2M} + K_{B1}\left[Ag^+\right]\mu_{[Ag(MXO)]} + K_{B1}K_{B2}\left[Ag^+\right]^2\mu_{[Ag_2(2M)]}}{1 + K_{B1}\left[Ag^+\right] + K_{B1}K_{B2}\left[Ag^+\right]^2}. \tag{12}$$

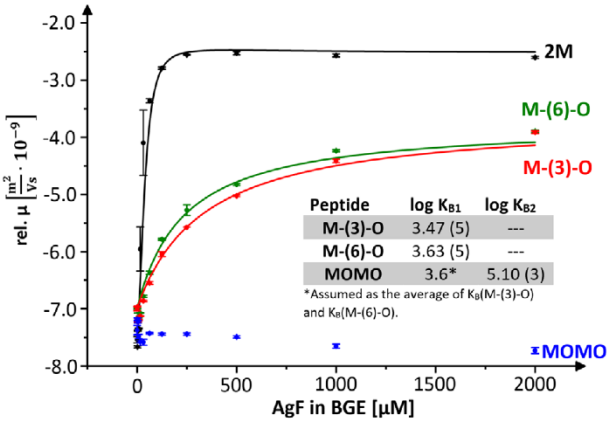


Figure 2. Mobility plot for the peptides 2 M (black); M-(3)-*O* (red); M-(6)-*O* (green); and MOMO (blue) in 50 mM phosphate buffer, pH 3. The corresponding lines are fits according to Eqs. (11) and (12). Displayed error bars represent standard deviations of four measurements. Separation voltage: 20 kV; capillary: fused silica. Inset table: Silver complex formation constants for [Ag(MXO)] and [Ag_2(2M)] according to Eqs. (9) and (10). Standard deviations of the fit are given in parentheses.

With these estimates, a very precise fit can be established which results log K_{B2}≈5.1 (Fig. 2).

Addition of 375 μM $AuCl_3$ to the phosphate BGE has a clear impact on 2 M, leading to a peak that has an increased mobility and a double peak with an unchanged mobility (Fig. 1B). The observed species with an increased mobility migrates in the same timeframe as the [Ag-2M] complex, thus, this signal is corresponding to a complex of Au ions and 2 M. Increasing concentrations of $AuCl_3$ in the BGE in order to shift the equilibrium to the peptide-gold-complex does not increase the intensity of the new peak. Instead, the newly formed peak first rises in intensity until approx. 500 μM $AuCl_3$, then intensity starts decreasing again (Fig. 1C). These data suggest that a more sophisticated mechanism is taking place, than a complex formation between the peptide and Au^{3+}. While a clear [Au-(MXO)] peak cannot be detected, peak fronting is detected for both single oxidized peptides at $AuCl_3$ concentrations up to 64 μM (Supporting information Figs. S1 and S2). With higher concentrations, a double peak is formed. This indicates that the [Au-(MXO)] complex is not as stable as [Ag-(MXO)].

The similarity of the migration of 2 M with $AuCl_3$ and AgF in the BGE, suggests the existence of Au(I) species in the BGE. Oxidation of Met with chloroauric acid ($HAuCl_4$) which can be formed in aqueous solutions out of $AuCl_3$ has been first reported by Bordignon et al. [40] and investigated in detail in more recent publications [41, 42]. The mechanism of the redox reaction is depicted in Fig. 3. In a first step, a [Au(III)–Met] complex (1) is formed quickly by a nucleophilic Cl^- substitution by the thioether of Met. This complex is rather unstable and can only be detected in traces in the first few minutes of the reaction. In a second step, an additional Met residue leads to the reduction of Au(III) to Au(I) [43]. As an intermediate product a chlorosulfonium-methionine ion is formed, which further hydrolyses to MetO [42]. The resulting [Au(I)–Met] complex (2) can dissociate to form free $[AuCl_2]^-$ or create a linear and polynuclear $[Au(I)\text{-}(Met)_2]_n$ complex (3). $[AuCl_2]^-$ is not stable in solution at room temperature but its elimination to solid Au(0) and $[AuCl_4]^-$ is reported to proceed slowly [44].

According to the position and number of free Met residues, the peptides investigated in this study show different migration changes in CZE when $AuCl_3$ is added to the BGE. The peptide containing two Met, 2 M, shows the most significant changes. In the context of an oxidation induced by $[AuCl_4]^-$, it is clear that there is a signal that corresponds to the oxidized MOMO peptide that was formed out of 2 M. MALDI measurements of an equimolar mixture of $AuCl_3$ and 2M peptide further confirmed that oxidation of 2 M to single oxidized and double oxidized species is induced by Au(III) (data not shown). [Au(III)–Met] is reported to be only detectable in traces by NMR analysis after 3 min of reaction, while $[Au(I)\text{–}(Met)_2]_n$ is stable for hours. This suggests that there is no direct detection of the [Au(III)–(2M)] complex taking place, but that the newly formed peak for 2 M is an [Au(I)–(2M)] species. The mobility of the newly formed peak also highly corresponds to the [Ag_2(2M)] complex (see Fig. 1), further sustaining this assumption. With low concentrations of $AuCl_3$ (2–16 μM) in the BGE, the 2 M peptide peak starts to show a small delay, without detection of the newly formed species (Fig. 1C). This indicates the complexation of 2M with Au(III), since this complex is not charged while it enlarges the hydrodynamic radius of the peptide. The small mobility change can also be detected for the single oxidized species, in a smaller range, but is not observed for the MOMO species, indicating that it is caused by complexation of thioether groups (Supporting information Figs. S1 and S2). With higher $AuCl_3$ concentrations, the second reaction step starts to play an increasing role, leading to the detection of [Au(I)-(2M)]. At a concentration of 250–500 μM of $AuCl_3$ in the BGE, this species is most stable for detection. An $[AuCl_4]^-$ excess like in the 1000–2000 μM lanes, leads to an increased reaction kinetics, so that the peak intensity of [Au(I)-(2M)] starts decreasing (Fig. 1C). Adjusting the pH of the BGE with HCl in order to increase concentration of Cl^-, leads to more $[AuCl_4]^-$ ions at the same $AuCl_3$ concentration. This change led to a reduced complex detection, indicating a faster reaction process (data not shown), which matches with observations of Glišić et al. [45]. It remains unclear, whether [Au(I)-(2M)] consists of one Au(I) bound to both sulfurs of the peptide, a dinuclear complex with Au(I) bridging between two peptides, or even more sophisticated polynuclear structures. NMR studies conducted on a Gly-Met dipeptide have revealed that this peptide forms polynuclear [Au(I)$(Gly\text{-}Met)_2$]n complexes [42]. Detection of an [$(Ag)_2$(2M)] complex, the closely related chemical properties of Au(I) and Ag(I) and the very similar mobility of [Au(I)(2M)] and [$(Ag)_2$(2M)] observed in our data (Fig. 1), suggests that two Au(I) ions bind to one 2 M peptide under separation conditions. This difference to literature might occur due to the special conditions within separation in comparison to NMR or as a consequence

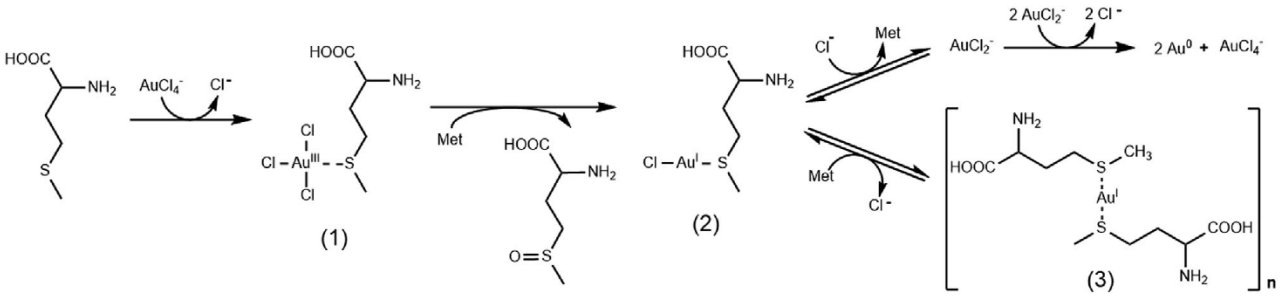


Figure 3. Reaction mechanism of Met with $AuCl_4^-$

of sterical problems resulting from the size of an 8-mer peptide.

4.2 Silver for the detection of oxidative stress in proteins

In a next step, the usability of the approach in a more sophisticated system was evaluated. For this, two mAbs were tested in a modified EACA buffer system with pH 5.7. This BGE has been demonstrated to have a broad applicability for biopharmaceuticals including ACE studies [31, 32, 46]. When it comes to the adaptation of the method to proteins, some aspects need to be considered. First, with a rising pH uncharged amino groups are available in the buffer, giving a new complexation partner for Ag(I). Although the pKa of the Met amino group is approximately 9, this effect is reported to be of relevance at pH > 4.5 [37].

Next, the affinity of silver toward free thiol groups is known to be higher than for thioethers, which means that Ag(I) may bind to Cys residues first. However, this effect should be constant for stressed and unstressed proteins, thus, resulting in similar mobility shifts for stressed and unstressed proteins. Additionally, most Cys residues are covalently bound in disulfide bonds, which were shown to have a lower affinity to Ag(I) than thioethers [35]. For our experiment, two standard mAbs, mAb1 and mAb2, each with a pI of 8.5 and with 14 Met residues were used as model biopharmaceuticals. These samples were stressed with 1% H_2O_2 for 1 or 6 h. For both mAbs, only minor CZE-profile changes are observed when they are exposed to H_2O_2, indicating increased oxidation (Fig. 4A and B). When these samples were injected in the presence of 500 μM AgF in the BGE, a mobility shift of unstressed/nonoxidized and intermediately (1 h) stressed samples occurred, while highly stressed samples (6 h) were not affected (Fig. 4C and D). This indicated binding of Ag(I) to free thioether groups and, thus, demonstrated, that stressed mAbs, containing MetO, were not affected by the presence of Ag(I) in the BGE while unstressed samples were changing their mobility (Fig. 4). Alhazmi and colleagues conducted several ACE studies on the effect of metal ions, including Ag(I), on proteins [47–49]. For these studies, a Tris(hydroxymethyl)-aminomethan (tris) BGE at pH 7.4 was used. Results were ambiguous for Ag: some proteins as human serum albumin was reported to have an increased mobility, while others are not affected or show mobility decrease. This can be explained by the complex formation of Ag at higher pH. In this case, the participation of amino groups in complex formation leads to polynuclear $[(Ag)_2(Met)_2]$ complexes [39]. Additionally, K_B for amino and thioether groups are reported to be in the same range, leading to a competitive situation. Since there are usually more accessible amino groups than thioether groups in a protein, this increased number of interactions may affect protein structure and, thus, mobility in very different ways. For these reasons, it is expected that the method presented here will show best results in terms of specificity at low pH (ideally < 4.5).

5 Concluding remarks

Methionine oxidation is linked to decreased stability of proteins and is, thus, of great interest as a critical quality attribute in pharmaceutical industry. For development of an easy and cost-efficient method for determination of the oxidation state in proteins, we apply ACE in combination with metal agents in the BGE. This approach has been tested with six different metals: Ni, Pd, Pt, Cu, Ag, and Au. The best results were obtained by using Au and Ag. Our data have shown that Au(III) is reduced to Au(I) and Met is oxidized to MetO in the capillary. Au(I) can then complex Met containing peptides and change their mobility in ACE. $AuCl_3$ can, thus, be used together with thioethers containing samples for studying behavior of Au(I) in CZE at concentration around 250–500 μM. Additionally, $AuCl_3$ can be of interest as an alternative oxidant for stressing biological samples.

The specific interaction of Ag(I) with thioethers could be successfully exploited in ACE to separate species with different oxidation status of their Met residues. Studies with peptides have shown a clear separation of nonoxidized, single-oxidized, and double- oxidized species, which could not be separated by classical CZE. Moreover, even different types of single-oxidized species were distinguished. Our data suggest the binding of one Ag(I) ion to one Met residue, which increases the overall charge of the analyte. The binding constant for 8-mer peptides containing one Met residue was determined to be log $K_B \approx 3.6$ and are well comparable to literature values obtained by amperometric titration.

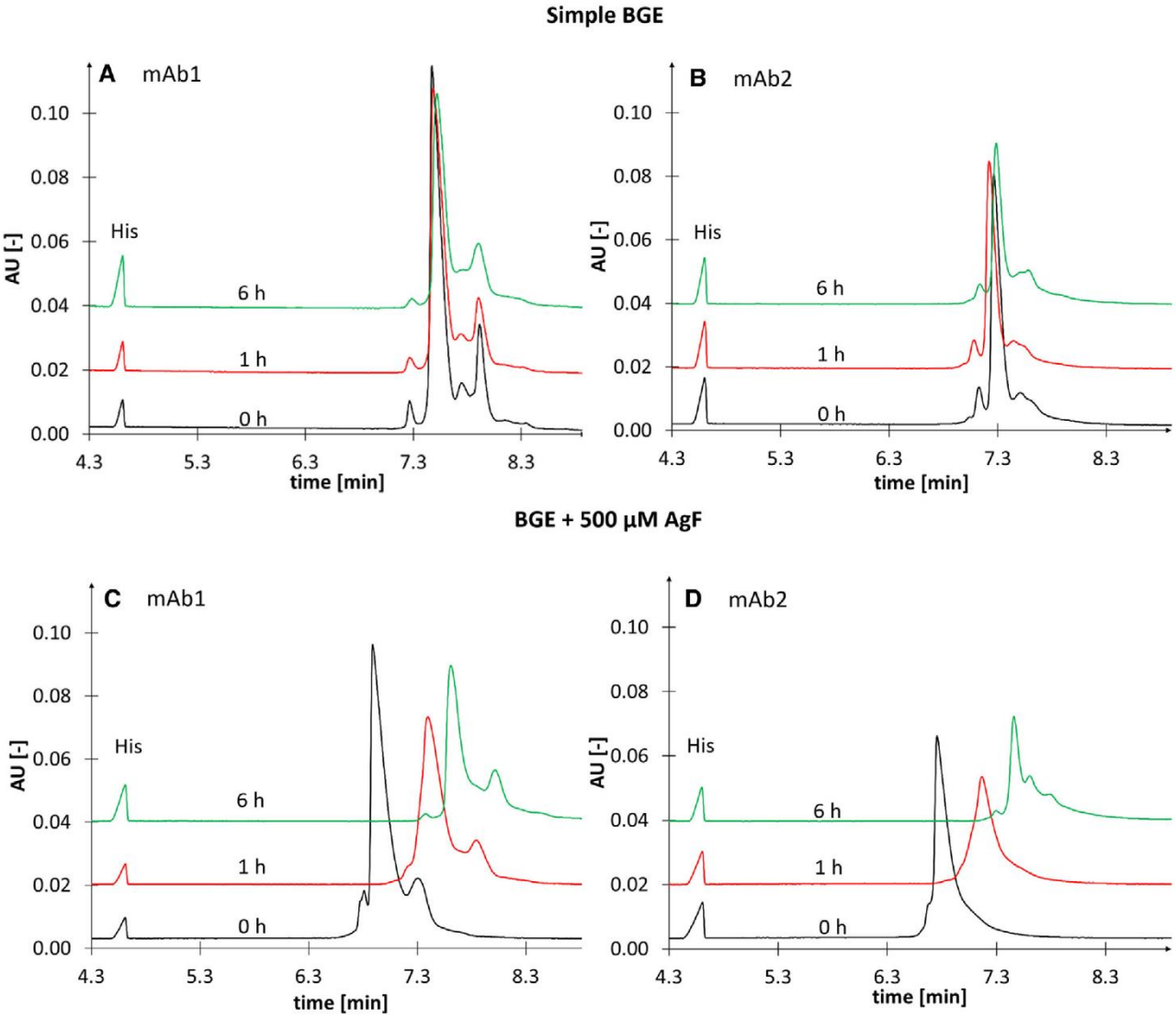


Figure 4. Effects of the AgF addition on the mobility of stressed mAb samples: mAb1 (A) and mAb2 (B) were stressed with 1% H_2O_2 for 1 h (red lines) and 6 h (green lines) and compared with unstressed samples (black lines) in a metal-free BGE. The same BGE was modified with 500 µM AgF and the experiment was repeated for mAb1 (C) and mAb2 (D). BGE: 400 mM EACA, 2 mM TETA, 0.05% HPMC, pH 5.7; separation voltage: 20 kV; capillary: fused silica. All lines were shifted according to the marker (His).

Experiments with mAbs have shown that the binding of silver ions also affects mobility of large proteins. However, mobility changes obtained for mAbs are relatively low compared to those of peptides, because of the low abundance of Met in the protein. Hence, for proteins the method has the ability to be used as an indicator for protein oxidation stress. If a more detailed resolution is necessary, the method needs to be used on peptide level.

The presented new ACE method for the detection of Met oxidation is economical and specific. Additionally, it is easily accessible as using standard equipment and techniques and has the advantage of rapid analysis times. It is expected to become a valuable tool for MetO detection and characterization.

The authors have declared no conflict of interest.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

6 References

[1] Cleland, J. L., Powell, M. F., Shire, S. J., *Crit. Rev. Ther. Drug Carrier Syst.* 1993, *10*, 307–377.

[2] Vogt, W., *Free Radic. Biol. Med.* 1995, *18*, 93–105.

[3] Bernstein, H. D., Poritz, M. A., Strub, K., Hoben, P. J., Brenner, S., Walter, P., *Nature* 1989, *340*, 482–486.

[4] Gellman, S. H., *Biochemistry* 1991, *30*, 6633–6636.

[5] Stadtman, E. R., Van Remmen, H., Richardson, A., Wehr, N. B., Levine, R. L., *Biochim. Biophys. Acta* 2005, *1703*, 135–140.

[6] Burney, P. R., White, N., Pfaendtner, J., *PLoS One* 2014, *9*, e86981.

[7] Sasaoki, K., Hiroshima, T., Kusumoto, S., Nishi, K., *Chem. Pharm. Bull.* 1989, *37*, 2160–2164.

[8] Frelinger, A. L., 3rd, Zull, J. E., *Arch. Biochem. Biophys.* 1986, *244*, 641–649.

[9] Teh, L. C., Murphy, L. J., Huq, N. L., Surus, A. S., Friesen, H. G., Lazarus, L., Chapman, G. E., *J. Biol. Chem.* 1987, *262*, 6472–6477.

[10] Mulinacci, F., Poirier, E., Capelle, M. A., Gurny, R., Arvinte, T., *Eur. J. Pharm. Biopharm.* 2013, *85*, 42–52.

[11] Stracke, J., Emrich, T., Rueger, P., Schlothauer, T., Kling, L., Knaupp, A., Hertenberger, H., Wolfert, A., Spick, C., Lau, W., Drabner, G., Reiff, U., Koll, H., Papadimitriou, A., *mAbs* 2014, *6*, 1229–1242.

[12] Tarrago, L., Peterfi, Z., Lee, B. C., Michel, T., Gladyshev, V. N., *Nat. Chem. Biol.* 2015, *11*, 332–338.

[13] Yan, B., Valliere-Douglass, J., Brady, L., Steen, S., Han, M., Pace, D., Elliott, S., Yates, Z., Han, Y., Balland, A., Wang, W., Pettit, D., *J. Chromatogr. A* 2007, *1164*, 153–161.

[14] Harris, R. J., *Dev. Biol.* 2005, *122*, 117–127.

[15] Lam, X. M., Yang, J. Y., Cleland, J. L., *J. Pharm. Sci.* 1997, *86*, 1250–1255.

[16] Anantharamaiah, G. M., Hughes, T. A., Iqbal, M., Gawish, A., Neame, P. J., Medley, M. F., Segrest, J. P., *J. Lipid Res.* 1988, *29*, 309–318.

[17] Boyd, D., Kaschak, T., Yan, B., *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* 2011, *879*, 955–960.

[18] Bouchon, B., Jaquinod, M., Klarskov, K., Trottein, F., Klein, M., Van Dorsselaer, A., Bischoff, R., Roitsch, C., *J. Chromatogr. B: Biomed. Appl.* 1994, *662*, 279–290.

[19] Ghesquiere, B., Gevaert, K., *Mass Spectrom. Rev.* 2014, *33*, 147–156.

[20] Nashabeh, W., Greve, K. F., Kirby, D., Foret, F., Karger, B. L., Reifsnyder, D. H., Builder, S. E., *Anal. Chem.* 1994, *66*, 2148–2154.

[21] Stettler, A. R., Chaurin, V., Constable, E. C., Housecroft, C. E., Schwarz, M. A., *J. Biol. Inorg. Chem.* 2007, *12*, 194–203.

[22] Stettler, A. R., Chaurin, V., Constable, E. C., Housecroft, C. E., Schwarz, M. A., *Electrophoresis* 2008, *29*, 3342–3348.

[23] Redweik, S., Xu, Y., Wätzig, H., *Electrophoresis* 2012, *33*, 3316–3322.

[24] Redweik, S., Cianciulli, C., Hara, M., Xu, Y., Wätzig, H., *Electrophoresis* 2013, *34*, 1812–1819.

[25] Pearson, R. G., *Surv. Prog. Chem.* 1969, *5*, 1–52.

[26] Kolthoff, J., Harris, W., *Ind. Eng. Chem. Anal. Ed.* 1946, *18*, 161–162.

[27] Benesch, R., Benesch, R. E., *Arch. Biochem.* 1948, *19*, 35–45.

[28] Bueton, H., *Biochim. Biophys. Acta* 1958, *29*, 193–201.

[29] Sun, R. W.-Y., Che, C.-M., *Coord. Chem. Rev.* 2009, *253*, 1682–1691.

[30] Neubert, R. H., Ruttinger, H.-H., *Affinity Capillary Electrophoresis in Pharmaceutics and Biopharmaceutics*, CRC Press, New York 2003.

[31] He, Y., Isele, C., Hou, W., Ruesch, M., *J. Sep. Sci.* 2011, *34*, 548–555.

[32] Hutanu, A., Kiessig, S., Bathke, A., Ketterer, R., Riner, S., Stracke, J. O., Wild, M., Moritz, B., *Electrophoresis* 2019, *40*, 3014–3022.

[33] McAuliffe, C., Murray, S., *Inorg. Chim. Acta Rev.* 1972, *6*, 103–121.

[34] Young, A. G., Hanton, L. R., *Coord. Chem. Rev.* 2008, *252*, 1346–1386.

[35] Sigel, H., Scheller, K. H., Rheinberger, V. M., Fischer, B. E., *J. Chem. Soc. Dalton Trans.* 1980, *7*, 1022–1028.

[36] Luehrs, D. C., Nicholas, R. W., Hamm, D. A., *J. Electroanal. Chem. Interfacial Electrochem.* 1971, *29*, 417–420.

[37] Pettit, L. D., Siddiqui, K. F., Kozłowski, H., Kowalik, T., *Inorg. Chim. Acta* 1981, *55*, 87–91.

[38] Jeżowska-Trzebiatowska, B., Kowalik, T., Kozłowski, H., *Bull. Acad. Pol. Sci. Ser. Sci. Chim.* 1977, *25*, 797–803.

[39] Lyons, A. Q., Pettit, L. D., *J. Chem. Soc. Dalton Trans.* 1984, *10*, 2305–2308.

[40] Bordignon, E., Cattalini, L., Natile, G., Scatturin, A., *J. Chem. Soc. Chem. Commun.* 1973,*22*, 878–879.

[41] Vujačić, A. V., Savić, J., Sovilj, S. P., Szécsényi, K. M., Todorović, N., Petković, M., Vasić, V. M., *Polyhedron* 2009, *28*, 593–599.

[42] Glišić, B. Đ., Rajković, S., Stanić, Z. D., Djuran, M. I., *Gold Bull.* 2011, *44*, 91–98.

[43] Natile, G., Bordignon, E., Cattalini, L., *Inorg. Chem.* 1976, *15*, 246–248.

[44] Gammons, C. H., Yu, Y., Williams-Jones, A., *Geochim. Cosmochim. Acta* 1997, *61*, 1971–1983.

[45] Glišić, B. Đ., Djuran, M. I., Stanić, Z. D., Rajković, S., *Gold Bull.* 2014, *47*, 33–40.

[46] Moritz, B., Schnaible, V., Kiessig, S., Heyne, A., Wild, M., Finkler, C., Christians, S., Mueller, K., Zhang, L., Furuya, K., *J. Chromatogr. B* 2015, *983*, 101–110.

[47] Alhazmi, H. A., Nachbar, M., Albishri, H. M., El-Hady, D. A., Redweik, S., El Deeb, S., Wätzig, H., *J. Pharm. Biomed. Anal.* 2015, *107*, 311–317.

[48] Alhazmi, H., Al Bratty, M., Javed, S., Lalitha, K., *Pharmazie* 2017, *72*, 243–248.

[49] Alhazmi, H. A., Javed, S. A., Ahsan, W., Rehman, Z., Al Bratty, M., El Deeb, S., Saleh, S. F., *Microchem. J.* 2019, *145*, 259–265.

SUPPORTING INFORMATION

Methionine oxidation of proteins analyzed by affinity capillary electrophoresis in presence of silver (I) and gold (III) ions

Electrophoresis. 2021 Jun; 42(11):1209-16.

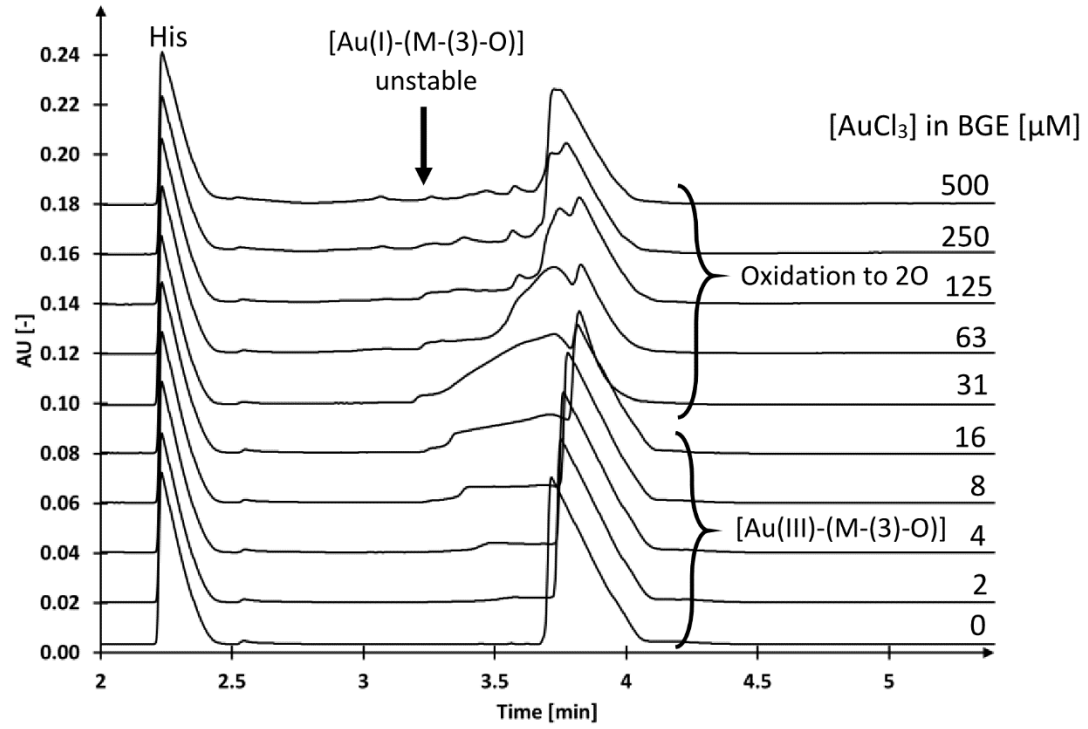


Fig. S1: Effects on the electropherogram of M-(3)-O with 0- 500 µM $AuCl_3$ in the BGE. All lines were adjusted according to the marker peak (His). Separation voltage of 20 kV; capillary: fused silica.

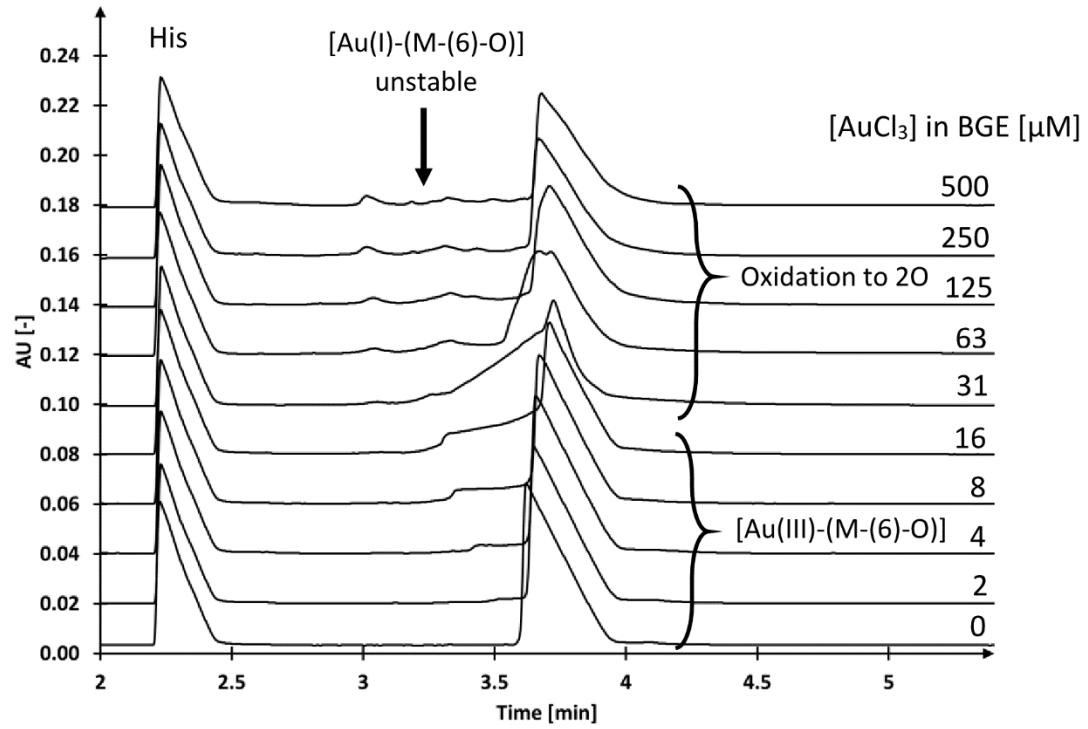


Fig. S2: Effects on the electropherogram of M-(6)-O with 0- 500 µM $AuCl_3$ in the BGE. All lines were adjusted according to the marker peak (His). Separation voltage of 20 kV; capillary: fused silica.

2.2 Analytical techniques for recombinant adeno associated virus

Recombinant adeno-associated viruses are gene therapy vectors, capable of transporting therapeutic genes to patient cells. [182] They are currently the most prevalent vector for in vivo gene therapy available to patients with three FDA approved drugs. [183] However, the production and quality control of rAAV poses significant challenges owing to the intricate and heterogenic nature of rAAV. The efficacy and safety of gene therapeutics are determined by the NA encapsulated, which, in turn, is associated with several CQAs. [184] Hence, the development of reliable and robust analytical methodologies for the characterization of rAAVs is of utmost significance. When this project was started, we realized that there are little to no reports that compare analytical methods with the same set of rAAVs. Through the large variability of commercial specimens and lack of reliable reference standards analytical method development was challenging. In order to overcome this problem a descriptive study was initiated to obtain a deeper understanding of the nature of rAAV samples. This chapter presents an evaluation of five analytical techniques for rAAV analysis, as described in two articles. The techniques evaluated include TEM, AUC, CGE, CZE and IP-RP-LC.

The first article takes a holistic approach and compares results obtained with TEM, AUC, CGE and CZE with a consistent set of rAAVs. TEM and AUC techniques are primarily concerned with the overall state of the viral system but also provide information on the DNA load. On the other hand, the described CE methods are suitable for analyzing the viral DNA specifically. TEM and AUC generated equivalent results, with no significant variations detected among the examined rAAVs: All species that were identified through TEM were also observable via AUC. AUC has the ability to offer a comprehensive assessment of the variability among sample species and was utilized to determine the proportional quantities of capsids that are empty, partially filled, filled, or overfilled. On the other hand, TEM is a powerful imaging technique that visualized impurities such as proteasomes or damaged virus capsids. The study revealed that the differentiation of partially filled from fully filled capsids through TEM and AUC techniques could pose a challenge. My contribution to this publication was to establish orthogonal CE techniques that could address this challenge. CGE provided a rapid estimation of the size and proportion of the present DNA molecules. The findings indicated that the detection of fragments in CGE resulted in the appearance of partially filled species in AUC. Capsids that encapsulate DNA strands larger than their intended size cannot be visualized through TEM and could thus only be observed through the use of CGE and AUC. The results obtained from CGE have demonstrated their comparability and complementarity to AUC. Although CGE proved to be a valuable tool, we observed that the

estimation of genome titer using this method was a laborious process and lacked an established calibration procedure. Hence, a simplified CZE workflow was proposed that can be utilized under native conditions with brief separation times (< 5 min/run). The underlying mechanism for detection was based on the release of DNA due to thermal stress, achieved by subjecting the sample to a temperature of 70 °C for a duration of 2 minutes. The experimental technique exhibited linearity over a wide range and yielded consistent outcomes. This approach could thus aid in comprehending the release of DNA, such as during formulation studies.

As a result of the accumulated knowledge, I was part of a collaboration with peers at the University of Leiden to assist in the development of an IP-RP-LC method that had not yet been implemented for rAAV DNA. The publication involved a comparative analysis of the acquired IP-RP data to AUC and CGE. I provided different preparations of rAAV and plasmid digests and the corresponding CGE data which helped highlighting the advantages and disadvantages of each technique. While CGE generally provides higher resolution than IP-RP-LC, the latter technique allows for analysis under fully denatured conditions, enabling detection of single-stranded DNA. Overall IP-RP-LC proved to be a valuable addition to AUC and CGE.

2ND PROJECT:

1st Publication

Stronger together: Analytical techniques for recombinant adeno associated virus

Electrophoresis. 2022 May; 43(9-10):1107-17.

Andrei Hutanu[1,2]
David Boelsterli[3]
Claudio Schmidli[3]
Cristina Montealegre[3]
Mike H. N. Dang Thai[1]
Balazs Bobaly[3]
Marius Koch[3]
Maria A. Schwarz[2,3]

[1]Analytical Development and Quality Control, Pharma Technical Development Biologics Europe, University of Basel, Basel 4056, Switzerland
[2]Analytical Development and Quality Control, Pharma Technical Development Biologics Europe, University of Basel, Basel 4056, Switzerland
[3]Solvias AG, Kaiseraugst, Switzerland

Received September 18, 2021
Revised October 25, 2021
Accepted November 13, 2021

Research Article

Stronger together: Analytical techniques for recombinant adeno associated virus

With recent FDA approval of two recombinant adeno-associated virus (rAAV)-based gene therapies, these vectors have proven that they are suitable to address monogenic diseases. However, rAAVs are relatively new modalities, and their production and therapy costs significantly exceed those of conventional biologics. Thus, significant efforts are made to improve the processes, methods, and techniques used in manufacturing and quality control (QC). Here, we evaluate transmission electron microscopy (TEM), analytical ultracentrifugation (AUC), and two modes of capillary electrophoresis (CE) for their ability to analyze the DNA encapsidated by rAAVs. While TEM and AUC are well-established methods for rAAV, capillary gel electrophoresis (CGE) has been just recently proposed for viral genome sizing. The data presented reflect that samples are very complex, with various DNA species incorporated in the virus, including small fragments as well as DNA that is larger than the targeted transgene. CGE provides a good insight in the filling of rAAVs, but the workflow is tedious and the method is not applicable for the determination of DNA titer, since a procedure for the absolute quantification (e.g., calibration) is not yet established. For estimating the genome titer, we propose a simplified capillary zone electrophoresis approach with minimal sample preparation and short separation times (<5 min/run). Our data show the benefits of using the four techniques combined, since each of them alone is prone to delivering ambiguous results. For this reason, a clear view of the rAAV interior can only be provided by using several analytical methods simultaneously.

Keywords:
Analytical ultra-centrifugation / Capillary gel electrophoresis / Capillary zone electrophoresis / Characterization / Transmission electron microscopy
DOI 10.1002/elps.202100302

Additional supporting information may be found online in the Supporting Information section at the end of the article.

1 Introduction

The adeno-associated virus (AAV) is a small, non-enveloped ssDNA virus of 25 nm in diameter that belongs to the family of parvoviridae [1]. The 4.7 kb ssDNA wild-type genome is surrounded by a capsid with icosahedral symmetry which is formed by three viral proteins (VP1–3) [2]. The capsid proteins are all coded by one gene (Cap) and are formed through alternative splicing and are thus only differing in their N-terminus. Additionally, the genome contains four different regulatory proteins (Rep) and is flanked by 145 nucleotide inverted terminal repeats (ITR), which form a T-shaped hairpin through base pairing [3]. The genome can either be a sense (+) or anti-sense (-) DNA strand [4]. Genome packaging into an empty capsid just needs ITRs, when all helper functions, Rep and Cap are provided by plasmids. In order to produce recombinant AAV (rAAV), 95% of the AAV genome can thus be removed and replaced by a transgene [5]. The two main production methods use transfection of mammalian Human Embryonic Kidney 293 (HEK293) cells or the baculovirus expression vector system (BEVS) for infecting *Spodoptera frugiperda* (Sf9) insect cells [6]. While the Sf9 system is easier to handle and shows better scalability, it is reported that rAAVs from this source tend to have a different expression of viral proteins, different

Correspondence: Andrei Hutanu, Analytical Development and Quality Control, Pharma Technical Development Biologics Europe, F. Hoffmann-La Roche AG, Grenzacherstrasse 124, 4070 Basel, Switzerland
E-mail: andrei.hutanu@roche.com

Abbreviations: **AEX**, anion exchange chromatography; **AUC**, analytical ultra-centrifugation; **CPA**, corrected peak area; **CQA**, critical quality attribute; **DBA**, dye based assay; **GC/mL**, genome copies per mL; **GMP**, good manufacturing practice; **HMW**, high molecular weight; **ITR**, inverted terminal repeat; **MWCO**, molecular weight cutoff; **QC**, quality control; **rAAV**, recombinant adeno associated virus; **S1 or S2**, supplier 1 or 2; **TEM**, transmission electron microscopy; **VP/mL**, virus particles per mL; **VP1–3**, virus proteins 1–3

posttranslational modifications, and different methylation of genomes, which may lead to a decreased infectivity [7,8].

The ability of site-specific integration of DNA into the host cell genome [9–11] and the missing pathogenicity [12] are attributes, that predestine AAVs to the use as gene delivery vehicles for gene therapy. With the FDA approval of two rAAV based gene therapies in 2017 and 2019 [13], these viruses have proven that they are suitable vectors to address monogenic diseases. In particular, rAAVs have shown remarkable improvements for patients with spinal muscular atrophy [14] and retinal diseases [15], but considerable research is also made for genetic remedies that affect the central nervous system, muscles, and liver [16]. Since rAAVs are relatively new modalities and production costs remain high, significant efforts are currently being made to improve the processes, methods, and techniques used in manufacturing and QC. The goal of analytical development for QC of biopharmaceuticals is to identify physicochemical properties that influence safety and efficacy and set specific ranges or limits that safeguard the high product quality that is required for therapeutic use. Most of these critical quality attributes (CQAs) have been identified for rAAVs as well and are discussed in recent reviews [17,18] (https://alliancerm.org/manufacturing/a-gene-2021). In this work we will focus on the CQAs discussed in the following section.

For appropriate dosing, the virus titer or concentration of the AAV sample is a product-related CQA. There are three general approaches to determine the titer. The genome titer (genome copies per mL: GC/mL) describes the capsids filled with the correct genome and is thus used for clinical dosing [19]. When the capsid concentration (viral particles per mL: VP/mL) is measured regardless of the content, the capsid titer is estimated. To get a measure of the biological activity, the infectivity titer is evaluated by different *in vitro* cell-based assays [20,21].

Production yields remain the main challenge for manufacturing rAAV because the majority of capsids tend to not contain the transgene [22]. For this reason, fast and automated methods for the estimation of the full to an empty ratio (or content ratio: full to all) are critical from a QC perspective. Capsids may also be filled with a part of the transduction cassette, the plasmid backbone, or wild-type sequences. This is expected to lead to immunotoxicity or genotoxicity [23].

Even though physicochemical methods are only rarely used in the characterization of viral vector systems during production, they are considered to play an important role in the future. Hereby, they will cover CQAs as the content ratio, the virus (protein) and genome titer, purity, or aggregation, to name just a few. TEM [24,25] and AUC [24,26] are two orthogonal techniques, that can detect different viral populations. Both have already been widely applied with their specific advantages and drawbacks. TEM is a very flexible method as it offers visual information for content ratio, protein impurities, and aggregation. However, inconsistent staining can lead to ambiguous results of capsid fillings and turnaround times are high. AUC, on the other hand, easily provides quantitative results for different viral species with a high resolution. Its limitations include high sample consumption and low throughput. Additionally, both methods seem to be challenging to implement in a good manufacturing practice (GMP) environment. Capillary gel electrophoresis (CGE) was recently proposed as a method that can separate different size variants of DNA in rAAV samples and could even have the prospect for genome titer estimation [27]. Related CGE methods are extensively used for the assessment of protein impurities of biopharmaceuticals including rAAVs [28–30]. In this publication, we discuss the advantage of the synergistic application of four methods, TEM, AUC, and two submodes of CE (see Table 1 [24,31–37]), focusing on the quantity, quality, and state of the genome. HPLC methods such as RP-HPLC, SEC or AEX, or other CE submethods as cIEF are not considered here. The applicability of these approaches and their place in the rAAV analytical toolbox is under investigation, although some reports consider these techniques already available [17,18]. The juxtaposition of these four methods revealed an in-depth understanding of the filling of complex rAAV samples and shed new light on the strengths and weaknesses of the CGE approach.

2 Materials and methods

General: HPLC grade water was prepared using a Q-POD® Ultrapure Water Remote Dispenser (Cat. no. ZMQSP0D01) by Merck Millipore/Merck KGaA (Darmstadt; Germany). Purified AAV2 and AAV3 (approx. 2×10^{13} GC/mL by qPCR) and the ssDNA genome were obtained from different commercial manufacturers. The incorporated genomes should express an enhanced green fluorescent protein (EGFP) with a Cytomegalovirus (CMV) promoter. Genome sizes have a theoretical size of 2501 bp for Supplier 1 and 2565 bp for Supplier 2. The samples from different sources are abbreviated with S1 (Supplier 1) and S2 (Supplier 2, e.g., AAV2 S2). Samples were formulated in 1x PBS with 0.001% Pluronic F-68.

TEM: Samples were stained with 1.5% (w/v) uranyl acetate in H_2O (Merck 8473; discontinued). TEM Grids: 200 mesh, copper, formvar/carbon-coated (Cat. no. S162) were from Plano GmbH (Wetzlar; Germany). Sample preparation: 4 μL of each sample was incubated for 30 s on a glow-discharged carbon-coated TEM grid. After a blotting step, grids were washed three times with Milli-Q water and negatively stained two times with 4 μL of 1.5% uranyl acetate. After every staining and washing step, excess liquid was removed using a blotting paper. Image acquisition was performed on a Philips CM100 TEM operated at 80 kV.

AUC: Analysis was performed on an Optima analytical ultracentrifuge from Beckman Coulter (Indianapolis, USA). The samples were loaded into AUC cells equipped with quartz windows and 12 mm double-sector charcoal-filled EPON centerpieces. A sample volume of 370 μL was loaded in the sample sector, whereas 390 μL of formulation buffer was loaded in the reference sector. The AUC cells were equilibrated at 15°C for 1 h before starting the rotation at the indicated rotor speed. The sedimentation was monitored

Table 1. Overview of Methods used in this work and the CQAs they address

Technique	CQA	Specific results	Comments	Requirements test sample	Quality standard	Possible Alternatives
TEM	Content ratio	Evaluation of empty, filled and disrupted AAV Evaluation of impurities	Uneven staining may lead to ambiguous results Simple visual interpretation	$>10^{12}$ VP/mL	non-GMP	AUC AEX [24,31] ELISA/qPCR [32] OD [32]
AUC	Content ratio	Estimation of filled and empty species Estimation of overfilled and partial filled species Estimation of degraded AAV	Empty material could be critical Re-homogenization and reutilization may be possible	$>10^{12}$ VP/mL ca. 400 μL Highly purified material No aggregates	non-GMP	TEM AEX [24,31] ELISA/qPCR [32] OD [32]
CGE_{DNA}	Quality of DNA Genome titer	Identity of genome (based on size/migration time) Size variants Quantification of target variant	Non-covalent labeling needed (BGE) CE separation profile depends on test sample preparation conditions	$>10^{11}$ GC/mL $>$ 0.2 ng/μL	GMP	AGE [19,33] AUC
CZE_{DNA}	Quantity of free DNA Genome titer	Specific DNA quantification based on q/r	Evaluation of encapsidated / free DNA Evaluation of DNA release under stress conditions	$>10^{11}$ GC/mL $>$ 0.2 ng/μL	GMP	qPCR [19,34] DBA [35–37]

AEX, anion exchange chromatography; qPCR, quantitative polymerase chain reaction; DBA, dye based assay

by UV absorbance at 230, 260, and 280 nm. Scans were collected every 60 s with a radial position data spacing of 10 μm. The sedimentation coefficient and relative amounts of each species were determined by processing the data with SEDFIT (v16.1c) using c(s) model (source1). The relative amount of each subpopulation of the AAV samples was acquired by integrating the respective peak area in the sedimentation plot.

CE: Amicon Ultra-0.5 mL Centrifugal Filters with a 100 kDa cutoff (Cat. no. UFC5100BK), Benzonase Nuclease (Cat. no. E1014), 0.5 M EDTA (Cat. no. 15 575 020), polyvinylpyrrolidone (PVP; Cat. no. 437 190), 10× Tris borate EDTA (TBE) buffer (Cat. no. 574 795), Tris Base (Cat. no T6066) and Urea (Cat. no. U6504) were supplied by Sigma–Aldrich/Merck KGaA (Darmstadt; Germany). DNase I 10× Reaction Buffer (Cat. no. AM8170G), PBS tablets (Cat. no. 18 912 014), Pluronic F-68 (Cat. no. 24 040 032), SYBR™ Green II RNA Gel Stain 10 000× concentrate in DMSO (Cat. no. S7564), UltraPure™ DNase/RNase-Free Distilled Water (Cat. no. 10 977 015) were purchased from Thermo Fisher Scientific (Waltham, USA). The QIAquick PCR purification Kit used for purification was from Qiagen (Germantown, USA). Acetic acid 99.7% (Cat. no. LC101001) was sourced from Fisher Chemical (Hampton; USA). Capillaries with an inner diameter of 50 μm (Cat. no. TSP-050375) and 100 μm (Cat. no. TSP-100375) were from Polymicro Technologies /Molex LLC (Phoenix, USA).

Analysis was carried out using a SCIEX PA800 Plus system (Brea; USA) equipped with a solid-state laser with an excitation wavelength of 488 nm and a 520 nm bandpass emission filter (Cat. no. 65–699) from Edmund Optics (Barrington; USA), a 30 kV power supply and a temperature-controlled autosampler (± 2°C). Data were acquired and analyzed using 32 Karat software 10.3. **CGE**: AAV samples received from the different manufacturers were purified following the instructions from the QIAquick PCR purification kit but with two washing steps of the QIAquick column. Before injection in CE, the sample was heated at 70°C for 2 min followed by 5 min in ice. The DNA digestion with benzonase and later centrifugation to remove benzonase and degraded material was performed as in [27]. After centrifugation, the sample was directly purified with the QIAquick PCR purification kit and heated at 70°C for 2 min followed by 5 min in ice. The separation gel buffer consisted of 1% PVP, 4 M Urea in 1xTBE solution with 1:25000 diluted SYBR Green II [27,38]. A bare fused silica capillary with a 100 or 50 μm internal diameter and 30 cm effective length was used for the separation. The separation voltage was 6 kV using reverse polarity. The samples were injected by applying –4 kV during 2–6 s. Capillary temperature was set to 25°C and 10°C was used for the autosampler. **CZE**: The separation buffer was prepared by dissolving Tris base to a final concentration of 25 mM and adjusting the pH to 8.0 ± 0.05 with acetic acid. SYBR-Green II was added to a final dilution of 1:10000 shortly before the analysis run. Purified linear construct which was diluted to the desired concentration with Nuclease-free distilled water served as a positive control, while formulation buffer served as blank. The same procedure as for the AAV sample was followed. Normal heat-shock: for CZE analysis 10 μL AAV sample was heated to 70°C for 2 min and placed on ice for 5 min immediately after. Variations in heating temperature and duration are described in a later section. Benzonase digest: to determine DNA content inside and outside capsids, the benzonase digest was performed as in the CGE with subsequent removal of the DNAse by centrifugal filters. The residue was collected

and heated to 70°C for 2 min and afterward placed on ice for 5 min. Analysis: a neutral capillary (Cat. no. 477 441) from SCIEX (Brea; USA) with an I.D. of 50 μm and cut to a total length of 30 cm and 20 cm length to window was employed. The capillary cartridge was kept at 25°C while samples were stored at 10°C. Prior to each injection, the capillary was rinsed with water and equilibrated with separation buffer at 50.0 psi for 1 min each. Samples were injected at 0.5 psi for 10 s and separated for 4 min with the separation voltage set to –30 kV.

After each sequence, the capillary was rinsed with water for 10 min at 50 psi. For long-term storage, the capillary was kept in the refrigerator with the capillary tips placed in water. Each capillary was used for several hundred injections without loss of performance.

3 Results and discussion

In the following sections, we evaluate TEM, AUC, and two modes of CE for their ability to analyze the DNA encapsidated by rAAVs. The methods discussed below are based on generally different separation and/or detection principles but also the "target molecule" is different. For TEM and AUC analysis, the intact viral systems are investigated whereas for CE methods the genome is set on focus. For this purpose, the virus has to be destroyed with suitable methods or the conditions have to be chosen in such a way that DNA can escape from the virus. Starting from the TEM and bridging to the AUC results, we take a deeper look into the viral filling, which to a significant extent, changes the physicochemical properties of the virus itself. Not only the weight (sedimentation coefficient), but also the charge inside and outside, and the zeta potential to name a few properties.

3.1 Transmission electron microscopy

Transmission electron microscopy (TEM) is a well-established method for providing visual information on AAVs. This allows an image-based morphological assessment of the capsid content [39,40]. The structural characterization relies on good contrast, which is typically achieved with heavy metal staining salts. The salt diffuses into the vacancies of the empty AAV capsids, and due to the electron-dense staining material, the core of the capsid appears dark in the TEM images. The higher the filling level (amount of encapsidated DNA), the less space within the capsid is available for the staining material, therefore the brighter the core regions of the AAVs appear (Fig. 1A). It should be noted, that in spite of the simple principle, automatic full/empty identification of AAV particles is still challenging, especially for partially filled particles. Novel software solutions use deep learning algorithms for quantitative assessments (e.g., from Vironova: https://cellculturedish.com/automated-integrity-analysis-of-aav-and-adenovirus-particles-using-minitem/). These approaches provide quite accurate results. Furthermore, TEM images reveal potential aggregates, impurities such as cell debris (Fig. 1C) broken particles (Fig. 1D), residual DNA (Fig. 1E). As one of the most abundant proteins with a size >500 kDa the 20S proteasome core is a common contamination that can easily be seen with TEM (Fig. 1F).

3.2 Analytical ultracentrifugation

As in a common centrifuge, species are separated in an AUC experiment by exposure to a centrifugal force. Depending on their size and weight, species sediment at a specific velocity through the AUC cell. Empty, full, partially filled and overfilled rAAV capsids can in this way be separated from each other. The architecture of the AUC allows furthermore the monitoring of the separation in real-time using, for example, absorbance detection [41]. Analysis of the data yields the intensity of the individual species as a function of their sedimentation coefficient, providing a sedimentation plot (Fig. 2A). The intensity is related to the extinction coefficients of the AAV capsid and the DNA, which both have a profound absorbance in the UV range (DNA maximum at 260 nm and capsid maximum at 280 nm). AUC is therefore mainly used to shed light on the content ratio as a CQA by analyzing at these wavelengths. As indicated before, the manufacturing of AAVs is still challenging and solutions often contain mixtures of species.

As a typical example, the sedimentation plot of the full AAV3 S2 shows five distinct signals of different intensity (Fig. 2A, line 1). The most intensive signal is detected with a sedimentation coefficient of 70 Svedberg. The relative standard deviation from six AUC measurements of this Svedberg value accounts to 0.1%. It is assigned to the full AAV3 species, in line with the TEM image that mostly shows this species (Fig. 1A). Furthermore, the intensity at 260 nm is larger than at 280 nm (Ratio 260/280 is approx. 1.24) indicating that the capsid contains DNA and is not empty. This argument is based on the fact that the capsid alone has its maximum absorbance at 280 nm (Fig. S1) and lower absorbance at 260 nm (Fig. 2A). By definition, the 260/280 ratio is therefore below 1 when looking at protein samples alone such as empty AAV capsids. On the contrary, the ratio for DNA is above 1 since its absorbance at 260 nm is larger than at 280 nm. A ratio of 1.24 for an AAV species in AUC can thus only stem from DNA-loaded capsids. The second most intense signal is detected at around 80 Svedberg, which means that it contains even more DNA than the full species. It is therefore assigned to an overfilled capsid as well as the signal between 90 and 100 Svedberg. Signals of smaller intensity can be detected below the full species of 70 Svedberg, indicating that they contain less or no DNA. A comparison to an empty AAV3 S2 batch shows a pronounced intensity at 50 Svedberg that is consequently associated with the empty capsid (Fig. 2A, line 2). Furthermore, the intensity ratio between 260 nm and 280 nm is 0.58, that is, a range expected for an empty capsid [41]. This also shows that 260/280 ratios above approximately 0.6 (and not only above 1) indicate that the corresponding capsids contain traces of DNA, that is, they are partially filled. This is the case

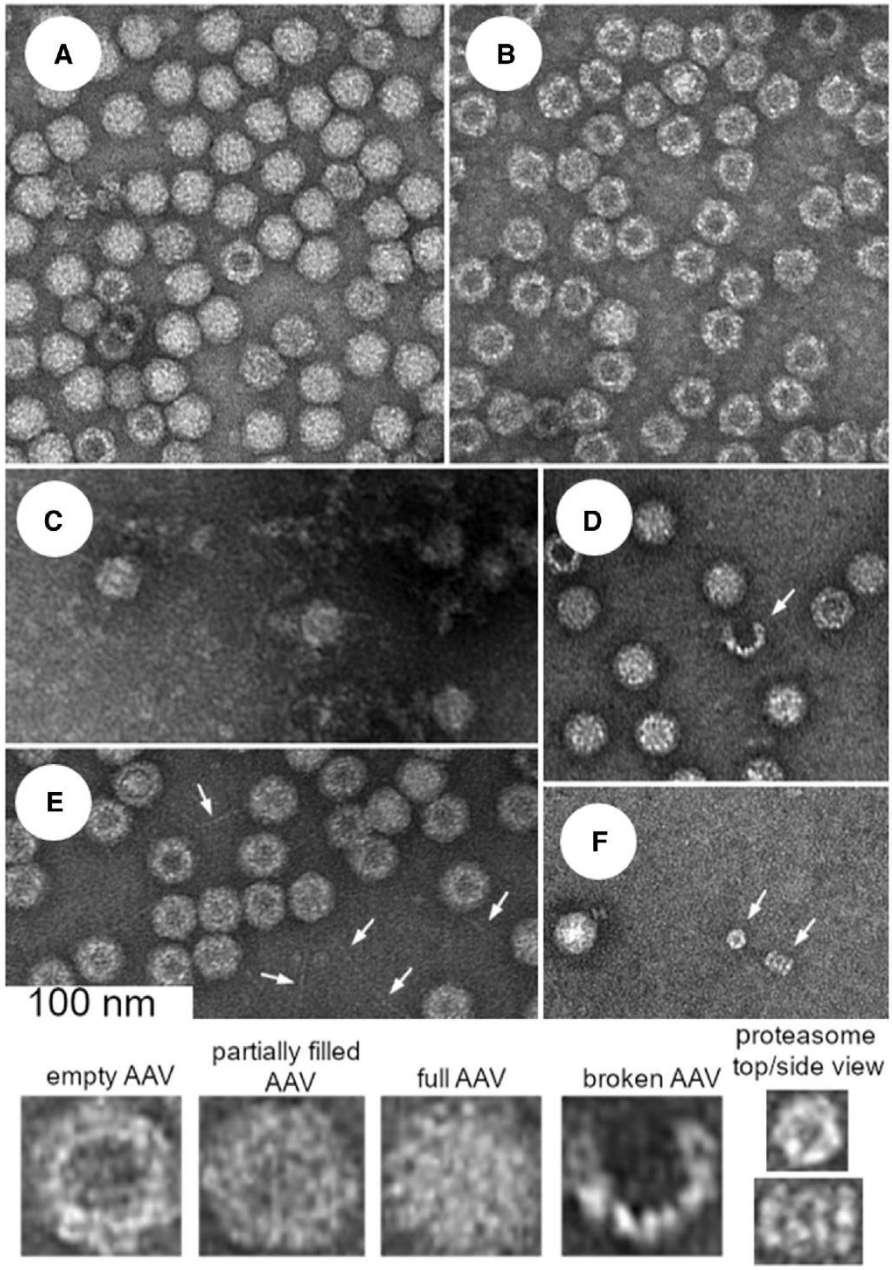


Figure 1. TEM images of AAV samples (upper part) with enlarged excerpts (lower part). (A) AAV3 S2 full, showing capsids with and without filling. (B) AAV3 S2 empty, showing capsids with and without filling. (C) AAV2 S1 full, showing impurities and aggregates in the background. (D) AAV2 S1 showing a broken AAV, (E) AAV3 S2 full, stressed, showing DNA in the background, (F) AAV2 S2 empty, after AUC showing residual proteasome impurities in the background.

for the signal at around 60 Svedberg, which falls in between empty and full capsids. The 260/280 ratio accounts to 1.22 for this signal.

Similar to the full AAV3 S2 batch, the empty AAV3 S2 batch showed five signals at 50, 60, 70, 80, and 90–100 Svedberg, but with different relative intensities. Here, the most intensive signal has already been assigned to the empty AAV3 and the remaining signal assignment follows the argumentation as for the full AAV3 S2 batch. It should be noted, that the AUC measurements of the AAV2 serotype were not repeatable and the samples degraded during the AUC run. Reuse of the solution was thus not possible and neither was an interpretation of the collected data.

To further ensure these results, the full and empty AAV3 S2 samples were mixed in a ratio of 1:1 and also analyzed. A separation of the empty and full capsids was visible (Fig 2A). The absorbance signal of the raw data consisted of an OD of 0.4 for both wavelengths at 280 and 260 nm. The theoretical relative content of empty and full capsids should be therefore approximately 50%.

AUC experiments were also conducted with a thermally stressed full AAV3 S2 batch (stressed at 70°C for 5 min). Slower and faster sedimenting species as in unstressed samples were present at 34 Svedberg and over 200 Svedberg respectively (Fig. 2B). We propose that the signal for the newly appearing but slower sedimenting species is caused by

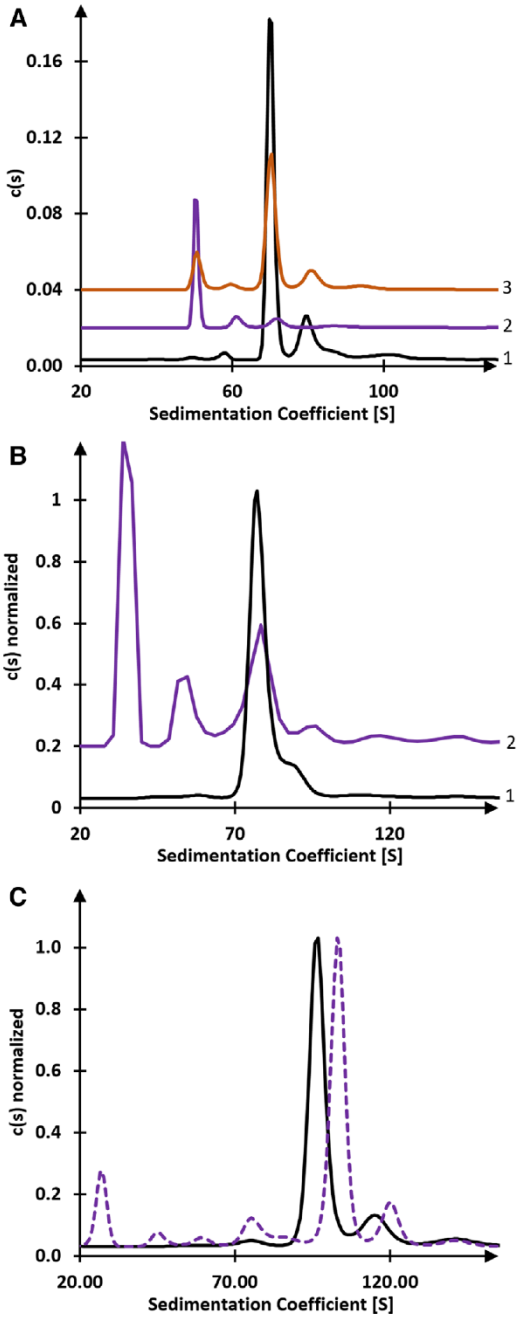


Figure 2. (A) Sedimentation distribution plot of AAV3 S2 showing the full batch (line 1), the empty batch (line 2), and the 1:1 mixture (line 3). The rotor speed was set to 20 000 rpm at 15°C and detection was conducted at 260 nm. Full and empty batches were diluted to a concentration of 7.80×10^{12} GC/mL based on the declared titer. (B) Normalized sedimentation distribution plot of AAV3 S2 showing the unstressed full batch (line 1) and the stressed full batch (line 2). Thermal stressing was carried out at 70°C for 5 min, the rotor speed was set to 16 000 rpm at 20°C. Detection was conducted at 260 nm with a sample concentration of 2.80×10^{12} GC/mL. (C) Normalized and corrected to standard conditions (water at 20°C) sedimentation distribution plot comparing full AAV3 batches from supplier 1 (dashed purple) and 2 (solid black). The rotor speed was set to 16 000 rpm at 20°C. The detection was conducted at 260 nm with a sample concentration of 2.80×10^{12} GC/mL.

capsid-fragments (34 Svedberg) and empty capsids (50 Svedberg) as detected in the stressed TEM samples (Fig. 1E). The faster sedimenting species are most probably aggregates. Furthermore, the intensity of the full capsid signal decreased whereas the signal for the empty capsid increased (Fig. 2B). The change indicates that the DNA diffuses/flows out of the capsid during thermal stressing (DNA release), which is also shown by TEM (Fig. 1E) and this can be exploited for genome quantification via CZE (see Fig. 4 later).

The collected data of the AAV3 serotype were compared between suppliers one and two. There is a slight shift in the sedimentation rate visible of the fully packed virus (Fig. 2C). This could be caused either by different packed DNA size or a general difference in the shape/weight of the vector. However, no significant differences of the packed DNA between the two distributors would be expected. By analyzing the content of the virus proteins VP1, VP2, and VP3 via CE-SDS, it was observed that the distributor with the faster sedimenting species showed a higher ratio of VP1 (not shown). This could explain the difference in the sedimentation behavior due to the fact that the VP1 protein has a higher molecular weight, compared to the VP2 and VP3 proteins.

3.3 Capillary gel electrophoresis

In the following, CGE data are used to support and complement the TEM and AUC results. With a CGE analysis, it is possible to separate all DNA size species, depending on their sequence length and conformation. In addition, it is generally possible to quantify by signal intensity using an appropriate reference standard. CGE data are complementary to qPCR and can be considered a useful extension in order to identify and characterize the genome, as CGE can distinguish between different product-related size variants, but also process-related variants as an oversized transgene or host cell DNA impurities.

The gel used in this analysis consisted of 1% PVP, 4 M Urea in 1×TBE buffer with 1:25000 diluted SYBR Green II [27,38]. While the electric field strength is an important parameter to consider in DNA analysis due to its high impact on separation efficiency and resolution, the optimum strength is mostly dependent on DNA size [42]. A low electric field of 150 V/cm together with 25°C capillary temperature were applied to have good resolution for a higher size range in around 30 min separation time. In addition, intermolecular conformation and intermolecular interaction can have a considerable impact on the separation profile. Thus, denaturing conditions during separation (4 M Urea in BGE) are indispensable [43]. Under these conditions, the separation of nine RNA transcripts from the 0.2 to 10 kb Transcript RNA Marker was possible with a resolution of 200 bases in the range of AAV genome (2 to 5 kb), although non-linear resolution is expected in this range [44].

The full AAV3 S1 and AAV3 S2 samples showed a similar profile with the main peak as the 2.5 kb ssDNA (Fig. 3 lines 2–4), which was also confirmed by spiking experiments

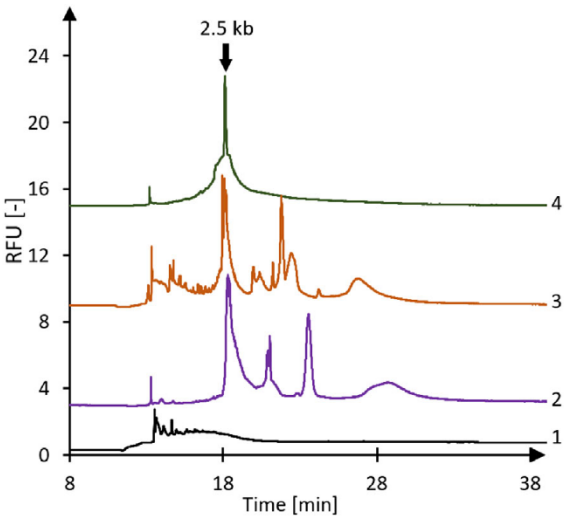


Figure 3. Comparison of CGE-LIF separation of (1) empty AAV3 S2, (2) full AAV3 S2, (3) full AAV2 S1 and (4) ssDNA genome from S1. Conditions for all lines: BGE: 1% PVP, 4 M Urea in 1×TBE with 1:25 000 SYBR Green II, Detection at 520 nm, Separation: 30/40 cm fs capillary; –6 kV at 25°C capillary temperature. Samples were purified with the QIAquick PCR purification kit. For further conditions, see chapter 2.

and migration time comparison. Larger impurities and a low amount of smaller DNA fragments can also be found in both samples. As Fig. 3 shows, empty AAV3 samples did not show any clear genome material in comparison with the respective full samples, confirming that these viral vectors do not contain the target transgene.

Benzonase was used to determine if the nucleic acid impurities observed in Fig. 3 are encapsidated or present outside the capsid, probably originated from host-cell or plasmid DNA. To initially check the performance of the benzonase treatment and subsequent filtration, the ssDNA material was submitted to this process with the later purification step with the QIAquick kit. No peak was observed in the sample treated with benzonase, which confirms the applicability of the conditions used (data not shown). The same samples from Fig. 3 were also tested by using these two preparations (with/without benzonase treatment) but no significant changes were observed. This allowed confirming that the observed peaks are only caused by encapsidated DNA.

Although comparable profiles were always obtained for all samples, variations in peak shape and relative intensity of the larger variants were observed between different sample preparations. In order to prove, that the observed peak profile is not caused by dsDNA as a method artifact, strategies were tested to reduce or control the variants, as the addition of 10% (v/v) formamide, 10% (v/v) DMSO, or 5 min sample sonication prior to the heating step at 70°C. Only formamide introduced a reduction of high molecular weight (HMW) peaks (Fig. S2). Thus, we assess that the observed peaks are related to different encapsulated ssDNA species.

When comparing CGE data with AUC and TEM some interesting observations can be made. The AAV3 S2 empty sample showed some full capsids in TEM (Fig. 1B) and two filled species in AUC (Fig. 2B). CGE also shows that there is DNA inside these samples, which contains smaller fragments than the target genome. These small-sized DNA fragments can also be seen in the full samples from both suppliers, although to a different extent. Most likely they are assessed as correctly filled in TEM and can be determined as partially filled by AUC, but may also add up as correctly filled in this technique. We assume that these capsids contain ssDNA that incorporates some hairpin, mimicking the ITR structure. Interestingly, samples from supplier two contained more of the low molecular weight DNA forms (Fig. 3 line 3) and also showed many partially filled species in AUC (Fig. 2C). Samples from both manufacturers also contain ssDNA that is larger than 2.5 kb, which is generally possible since the AAV capsid can incorporate up to 5 kb [45]. A common wrong filling takes place when the plasmid backbone including the ITRs is packaged inside the virus [46]. We would expect a plasmid backbone size of approx. 2.8 kb for Supplier 1 and 4.5 kb for Supplier 2. These differences can also be seen in different migration times of the HMW DNA in Fig. 3 (lines 2 and 3), which seems to fit this theory. An approximate look at the CPAs from Fig. 3 suggests that more than 30% of the viral capsids are filled with DNA larger than intended. Although one additional species can be expected due to the packaging of the plasmid backbone, both suppliers show at least two DNA species that are inside the virus. In AUC heavy species are detected too (80 and 90 and 100 Svedberg Fig. 2A) but the fraction seems to be smaller than CGE suggests. It is thus very likely that these AUC species are caused by the incorporation of a wrong DNA. Besides the capsid backbone, it remains unclear what these DNA species may be.

3.4 Capillary zone electrophoresis

The charge density in oligonucleotides is uniform and constant for nucleic acid molecules of varying lengths. Therefore, they typically migrate based on the charge to hydrodynamic radius ratio at the same velocity within an electrical field in a free solution. Thus, the separation of nucleic acid species, according to their number of nucleotides or nucleotide sequence, requires either the addition of a sieving matrix (see CGE) or the addition of nucleic acid species to the BGE that can separate the ssDNA in a sequence-selective manner (affinity capillary electrophoresis). Otherwise, the signal intensity can be evaluated in order to get information on the total genome concentration. Here, we start with the discussion on how a simple CZE could be used in order to quantify nucleic acid species getting a result that correlates to the genome titer. However, it should be noted that this type of analytics is in its infancy and still needs many investigations, both CZE with and without additives in the BGE.

As already mentioned, the workflow for CGE showed a poor reproducibility in signal intensity, which most likely is

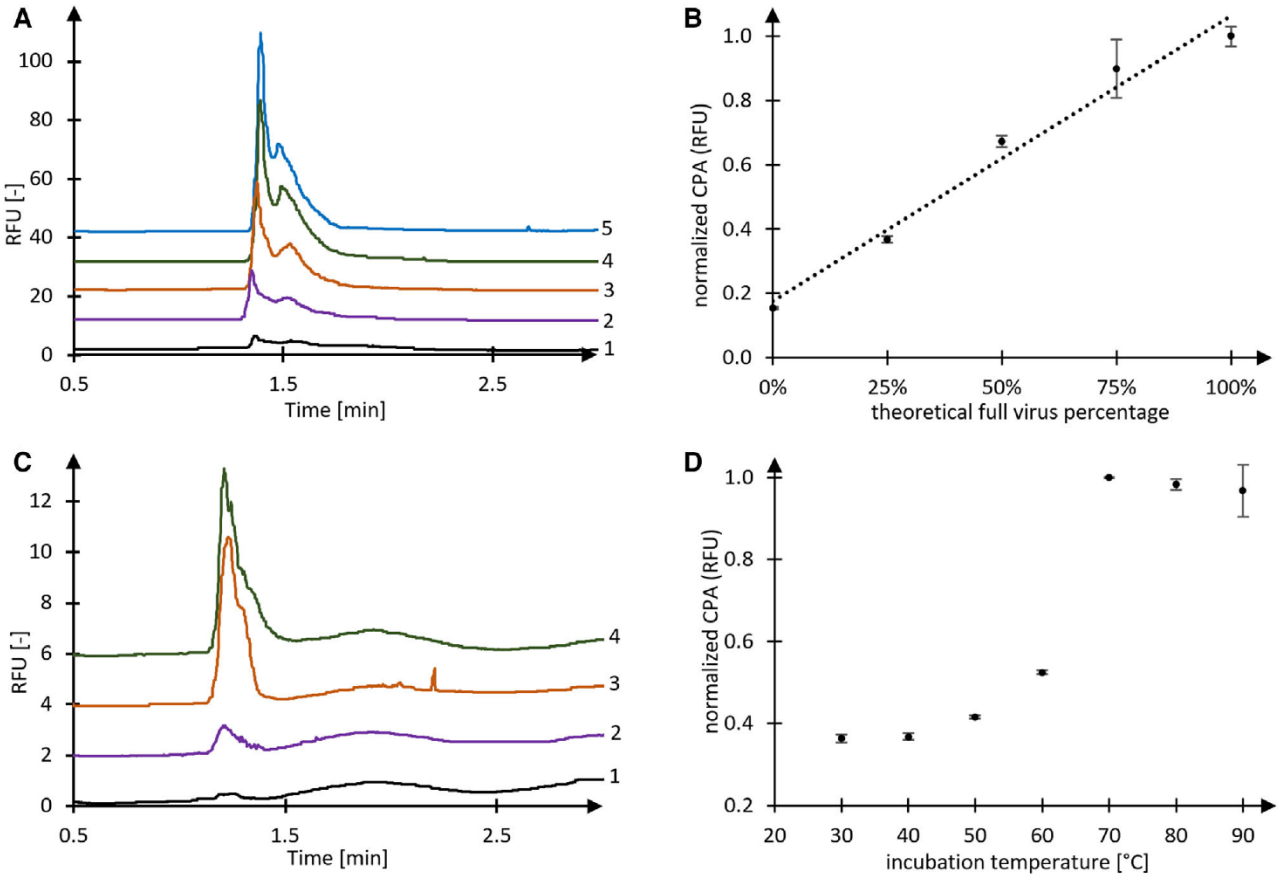


Figure 4. (A) Electropherogram of different mixtures of full and empty AAV2, S1 samples. Total concentration of each sample: 10^{13} VP/mL. Line 1: 0% full; Line 2: 25% full; Line 3: 50% full; Line 4: 75% full; Line 5 100% full. (B) Linearity of determined CPAs from A) Error bars represent standard deviation from four measurements. CPAs were normalized by setting the highest CPA (line 5 in A) as reference. (C) Electropherograms of differently treated full AAV2, S1 samples. Line 1: benzonase; no heat stress; no DNA Line 2: no benzonase; no heat stress; DNA impurities; Line 3: benzonase; heat stress; released encapsidated DNA Line 4: no benzonase; heat stress; released encapsidated DNA + impurities. (D) CPAs of differently stressed full AAV2, S1 samples. Stress conditions: 2 min of indicated temperature. Error bars represent standard deviation from two measurements. CPAs were normalized by setting the highest CPA (70°C measurement) as reference. Conditions for all measurements: BGE: Tris 25 mM; SYBR Green II 1:10 000; Detection at 520 nm, Separation: 20/30 cm neutral coated capillary; –30 kV at 25°C capillary temperature.

caused by the two purification steps. CZE could be a good alternative to determine the DNA titer, whereby the test sample preparation is clearly simpler and no purification steps are needed. The analysis is based on the release of the DNA after rupture of full capsid forced by short thermal stress (70°C, 2 min). Furthermore, the analysis of free DNA which is already outside of the virus system, or DNA leaked after specific stress conditions could be an interesting tool for the characterization and evaluation of the status of the virus shell.

As a proof of concept, we have diluted an AAV full sample with a formulation buffer (not shown) or mixed with an empty probe, and determined the linearity of the total peak area (Fig. 4A). For the analysis of the signal intensity and thus the concentration of DNA, the entire peak area was evaluated. Using a very simple BGE consisting of 25 mM Tris and 1:10000 SYBR Green II, a good correlation of virus concentration and the fluorescence signal was observed in the range of 10^{11} and 10^{13} GC/mL ($R^2 = 0.98$). The mixture with empty virus has a stable viral particle count but different DNA concentration and indicates that the method could also be used for estimation of the content ratio when combined with a capsid protein/particle measurement technique. We have found that the CZE profile, showing a sharp signal followed by a broader one, strongly depends on the salt concentration inside the sample. Experiments conducted with pure DNA diluted in different concentrations of PBS have shown, that at low salt concentrations only one peak is obtained, while a rising concentration leads to a more complex profile due to binding salt ions (not shown). To ensure that the detected peaks were indeed DNA, we have performed a benzonase digest after temperature stress which was able to remove all observed peaks.

For further assessment of applications of the CZE method, we have performed a similar workflow as for CGE,

with the aim to detect the DNA fraction located outside of the virus and/or the DNA inside the virus. For this purpose, the sample was either treated with benzonase with subsequent removal of the DNAse via centrifugal filters (100 kDa MWCO) or just filtered for comparison. Both preparations were split and one half was stressed for 2 min at 70°C and the other half remained untreated. Figure 4C shows that the method is indeed able to detect DNA outside of the capsid that is cut by benzonase (Lines 1 and 2) and is thus suitable to estimate DNA impurities that are not inside the capsids. Using suitable DNA reference standards, quantification could also be carried out, but it must first be thoroughly clarified which conditions allow stable DNA analysis.

Another interesting application for this assay would be all kinds of stability studies or studies related to storage conditions (e.g., freeze/thaw, impact of UV irradiation) or formulation development. This setup was used to ensure that unstressed viral samples are not leaking DNA when a strong electric field (1 kV/cm) is applied, by comparing the peak size of a DNA standard with rAAV sample peak size from 83 to 1000 V/cm. Our results showed no DNA release at any field strength, indicating that the method is indeed suitable for rAAV. Additionally, we have estimated the release of viral DNA through thermal stress for a temperature ranging from 30–90°C and stress duration of 2–20 min (electropherograms not shown). Fig. 4D shows that the critical temperature for DNA release is between 60 and 70°C, while little DNA is released below 60°C. This is in line with the results obtained in TEM (Fig. 1E) which showed free DNA and more empty particles (also demonstrated by AUC Fig. 2B) when the virus was stressed at 70°C. There was no crucial difference detected between different durations meaning that the process of DNA release is very fast and is completed within 2 min. Our data shows how this approach can be used to detect different conditions which lead to DNA release, possibly helping to understand the viral unpacking process.

4 Concluding remarks

Focusing on the quantity, quality, and state of the genome this work has shown the effectiveness of a synergistic approach of orthogonal physicochemical methods. While TEM and AUC set the focus on the state of the complete viral system, with respect to genome filling, the presented CE methods can be used for the analysis of the genome itself. TEM and AUC delivered comparable results, without major discrepancies between the viral entities observed. All species identified by TEM were detected by AUC as well. AUC is capable of providing an overall picture on the heterogeneity of the samples, which can be used for the determination of relative amounts of empty, partially filled, filled, and overfilled capsids. TEM generates impressive images and can be beneficially used for the visualization of impurities, like proteasomes or broken virus capsids. CE techniques can help to clarify results obtained with TEM and AUC, which alone would be difficult to interpret.

For example, partially filled capsids might be challenging to distinguish from correctly filled ones by TEM and AUC, as was demonstrated here for the samples that were supposed to be containing empty particles. In case the determination of the filling consistency by AUC or TEM remains challenging, CGE can quickly reveal an approximate size and relative amount of the encapsidated DNA strands. Our results showed that if fragments were identified in CGE, partially filled species always appeared in the AUC. Capsids containing larger DNA strands than intended cannot be detected by TEM at all, and are only seen by CGE and AUC. Although we would expect only one oversized DNA form, that would represent the plasmid backbone, AUC and CGE have revealed at least two additional species in samples from both suppliers. CGE has thus shown that it provides comparable and complementary results to AUC. In our opinion, it is conceivable that CGE will play an important role for the assessment of rAAV DNA purity in the future.

We found that CGE provides a good insight in the filling of rAAVs, but for estimation of the genome titer, the workflow is tedious and the process of calibration is not yet established. For this reason, we propose a simplified CZE workflow that can be used under native conditions with short separation times (< 5 min/run). The detection principle relies on DNA release under heat stress (2 min at 70°C). The method was shown to be linear in a broad range and provides reproducible results. Although dye based assays are using the same principle with 96 well-plate in a fluorescence plate reader have already been reported [35–37], the current CZE approach offers the benefit of selectively detecting the pure DNA, since it is very likely that DNA and viral proteins are separated under these conditions. This should lead to less artifacts and better reproducibility. DNA concentration could be determined more accurately and precisely, when using an appropriate reference standard, and this way support the understanding of DNA release under different conditions.

The complexity of rAAVs as next-level biopharmaceuticals is much higher compared to conventional protein biologics, e.g., monoclonal antibodies. Thus, relying on a single analytical technique increases the risk of observing ambiguous data or even misleading results in quality control. The four methods presented here offer a comprehensive analytical package to assess potential CQAs related to the capsid filling, and their use should therefore be considered in combination. Further developments and improvements, especially for CE-based methods are expected, since these approaches can be easily automated, offer fast analysis, and can be operated with low sample amounts.

The authors wish to thank for their contribution to this project Dr. Bernd Moritz for his useful and constructive suggestions during the planning and development of this research work, Dr. Adelheid Rohde, Dr. Manuel Gregoritza and Dr. Jan Stracke for their valuable support, and Gianni Morson for setting up our transmission electron microscope and establishing TEM as a technique at Solvias AG. Resources provided for this research by Solvias AG are gratefully acknowledged.

The authors have declared no conflict of interest.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

5 References

[1] Naso, M. F., Tomkowicz, B., Perry, W. L., Strohl, W. R., *BioDrugs* 2017, *31*, 317–334.

[2] Berns, K. I., Giraud, C., In: Berns, K. I., Giraud, C. (Eds.), *Adeno-Associated Virus (AAV) Vectors in Gene Therapy*, Springer, Berlin 1996 pp. 1–23.

[3] Lusby, E., Fife, K., Berns, K. I., *J. Virology* 1980, *34*, 402–409.

[4] Mayor, H. D., Torikai, K., Melnick, J. L., Mandel, M., *Science* 1969, *166*, 1280–1282.

[5] Samulski, R. J., Chang, L.-S., Shenk, T., *J. Virology* 1989, *63*, 3822–3828.

[6] Ayuso, E., Mingozzi, F., Bosch, F., *Curr. Gene. Ther.* 2010, *10*, 423–436.

[7] Rumachik, N. G., Malaker, S. A., Poweleit, N., Maynard, L. H., Adams, C. M., Leib, R. D., Cirolia, G., Thomas, D., Stamnes, S., Holt, K., *Mol. Ther. Methods Clin. Dev.* 2020, *18*, 98–118.

[8] Kondratov, O., Marsic, D., Crosson, S. M., Mendez-Gomez, H. R., Moskalenko, O., Mietzsch, M., Heilbronn, R., Allison, J. R., Green, K. B., Agbandje-McKenna, M., *Mol. Ther.* 2017, *25*, 2661–2675.

[9] Kotin, R. M., Siniscalco, M., Samulski, R. J., Zhu, X. D., Hunter, L., Laughlin, C. A., McLaughlin, S., Muzyczka, N., Rocchi, M., Berns, K. I., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 2211–2215.

[10] Daya, S., Cortez, N., Berns, K. I., *J. Virol.* 2009, *83*, 11655–11664.

[11] Wang, D., Tai, P. W., Gao, G., *Nat. Rev. Drug Discovery* 2019, *18*, 358–378.

[12] Boutin, S., Monteilhet, V., Veron, P., Leborgne, C., Benveniste, O., Montus, M. F., Masurier, C., *Hum. Gene. Ther.* 2010, *21*, 704–712.

[13] Kuzmin, D. A., Shutova, M. V., Johnston, N. R., Smith, O. P., Fedorin, V. V., Kukushkin, Y. S., van der Loo, J. C., Johnstone, E. C., *Nat. Rev. Drug Discovery* 2021, *20*, 173–174.

[14] Al-Zaidy, S., Pickard, A. S., Kotha, K., Alfano, L. N., Lowes, L., Paul, G., Church, K., Lehman, K., Sproule, D. M., Dabbous, O., *Pediatr. Pulmonol.* 2019, *54*, 179–185.

[15] Miraldi Utz, V., Coussa, R. G., Antaki, F., Traboulsi, E. I., *Ophthalmic Genet.* 2018, *39*, 671–677.

[16] Keeler, A. M., Flotte, T. R., *Annu. Rev. Virology* 2019, *6*, 601–621.

[17] Wright, J. F., *Biotechnol. J.* 2021, *16*, 2000022.

[18] Gimpel, A. L., Katsikis, G., Sha, S., Maloney, A. J., Hong, M. S., Nguyen, T. T., Wolfrum, J., Springs, S. L., Sinskey, A. J., Manalis, S. R., *Mol. Ther. Methods Clin. Dev.* 2021, *20*, 740–754.

[19] d'Costa, S., Blouin, V., Broucque, F., Penaud-Budloo, M., Fçranois, A., Perez, I. C., Le Bec, C., Moullier, P., Snyder, R. O., Ayuso, E., *Mol. Ther. Methods Clin. Dev.* 2016, *3*, 16019.

[20] Zhen, Z., Espinoza, Y., Bleu, T., Sommer, J. M., Wright, J. F., *Hum. Gene Ther.* 2004, *15*, 709–715.

[21] François, A., Bouzelha, M., Lecomte, E., Broucque, F., Penaud-Budloo, M., Adjali, O., Moullier, P., Blouin, V., Ayuso, E., *Mol. Ther. Methods Clin. Dev.* 2018, *10*, 223–236.

[22] Penaud-Budloo, M., François, A., Clément, N., Ayuso, E., *Mol. Ther. Methods Clin. Dev.* 2018, *8*, 166–180.

[23] Davé, U. P., Cornetta, K., *Mol. Ther.* 2021, *29*, 418–419.

[24] Fu, X., Chen, W.-C., Argento, C., Clarner, P., Bhatt, V., Dickerson, R., Bou-Assaf, G., Bakhshayeshi, M., Lu, X., Bergelson, S., *Hum. Gene Ther. Methods* 2019, *30*, 144–152.

[25] Horowitz, E. D., Rahman, K. S., Bower, B. D., Dismuke, D. J., Falvo, M. R., Griffith, J. D., Harvey, S. C., Asokan, A., *J. Virol.* 2013, *87*, 2994–3002.

[26] Burnham, B., Nass, S., Kong, E., Mattingly, M., Woodcock, D., Song, A., Wadsworth, S., Cheng, S. H., Scaria, A., O'Riordan, C. R., *Hum. Gene Ther. Methods* 2015, *26*, 228–242.

[27] Luo, J., Guttman, A., *Adv. Biopharm. Anal.* 2020, *33*, 33–38.

[28] Zhang, C. X., Meagher, M. M., *Anal. Chem.* 2017, *89*, 3285–3292.

[29] Zhang, Z., Park, J., Barrett, H., Dooley, S., Davies, C., Verhagen, M. F., *Hum. Gene Ther.* 2021, *32*, 628–637.

[30] Kurasawa, J. H., Park, A., Sowers, C. R., Halpin, R. A., Tovchigrechko, A., Dobson, C. L., Schmelzer, A. E., Gao, C., Wilson, S. D., Ikeda, Y., *Mol. Ther. Methods Clin. Dev.* 2020, *19*, 330–340.

[31] Qu, G., Bahr-Davidson, J., Prado, J., Tai, A., Cataniag, F., McDonnell, J., Zhou, J., Hauck, B., Luna, J., Sommer, J. M., *J. Virol. Methods* 2007, *140*, 183–192.

[32] Smith, P. H., Parthasarathy, S., Isaacs, J., Vijay, S., Kieran, J., Powell, S. K., McClelland, A., Wright, J. F., *Mol. Ther.* 2003, *7*, 122–128.

[33] Fagone, P., Wright, J. F., Nathwani, A. C., Nienhuis, A. W., Davidoff, A. M., Gray, J. T., *Hum. Gene Ther. Methods* 2012, *23*, 1–7.

[34] Wang, Y., Menon, N., Shen, S., Feschenko, M., Bergelson, S., *Mol. Ther. Methods Clin. Dev.* 2020, *19*, 341–346.

[35] Piedra, J., Ontiveros, M., Miravet, S., Penalva, C., Monfar, M., Chillon, M., *Hum. Gene Ther. Methods* 2015, *26*, 35–42.

[36] Xu, J., DeVries, S. H., Zhu, Y., *Hum. Gene Ther.* 2020, *31*, 1086–1099.

[37] Bee, J. S., O'Berry, K., Zhang, Y. Z., Phillippi, M. K., Kaushal, A., DePaz, R. A., Marshall, T., *J. Pharm. Sci.* 2021, *110*, 3183–3187.

[38] Khandurina, J., Chang, H.-S., Wanders, B., Guttman, A., *BioTechniques* 2002, *32*, 1226–1230.

[39] Chen, H., *Microsc. Microanal.* 2007, *13*, 384–389.

[40] Zhou, Z. H., *Structure* 2012, *20*, 1286–1288.

[41] Schuck, P., Zhao, H., Brautigam, C. A., Ghirlando, R., *Basic principles of analytical ultracentrifugation*, CRC Press, Boca Raton, FL 2016.

[42] Heller, C., *Electrophoresis* 2000, *21*, 593–602.

[43] Song, J. M., Yeung, E. S., *Electrophoresis* 2001, *22*, 748–754.

[44] Durney, B. C., Crihfield, C. L., Holland, L. A., *Anal. Bioanal. Chem.* 2015, *407*, 6923–6938.

[45] Wu, Z., Yang, H., Colosi, P., *Mol. Ther.* 2010, *18*, 80–86.

[46] Chadeuf, G., Ciron, C., Moullier, P., Salvetti, A., *Mol. Ther.* 2005, *12*, 744–753.

SUPPORTING INFORMATION

Stronger together: Analytical techniques for recombinant adeno associated virus

Electrophoresis. 2022 May; 43(9-10):1107-17.

Supporting information

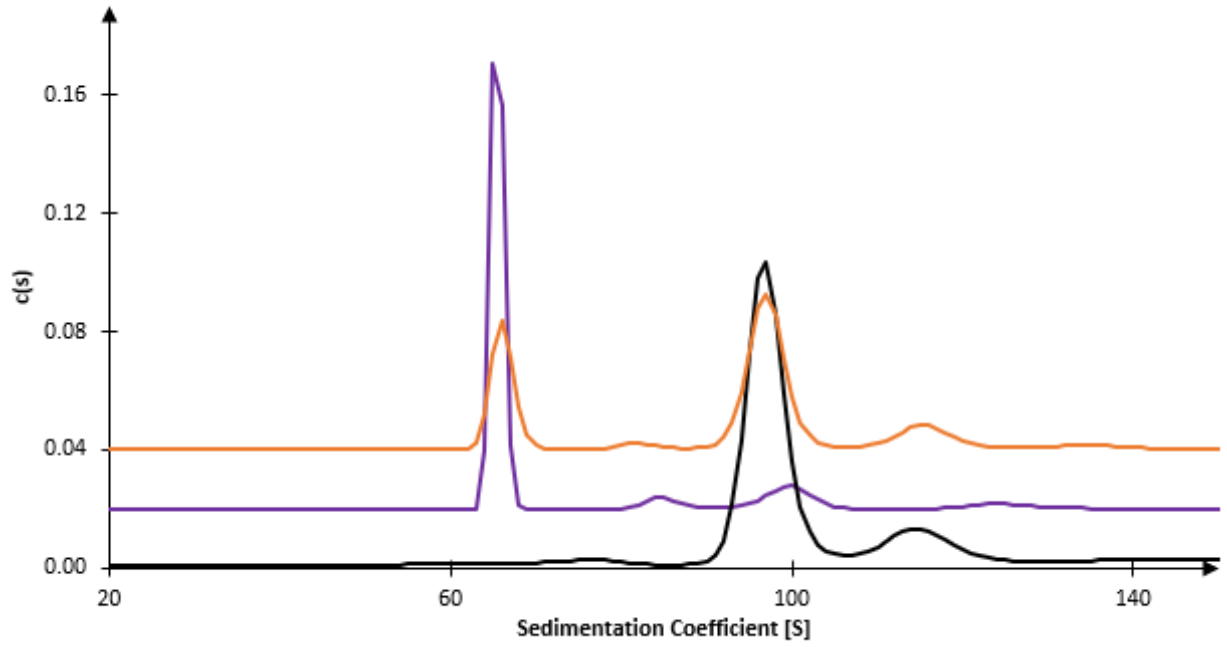


Fig. S1. Sedimentation distribution plot of AAV3 S2 showing the full batch (black), the empty batch (purple) and the 1:1 mixture (orange). The rotor speed was set to 20'000 rpm at 15 °C and detection was conducted at 280 nm. Full and empty batches were diluted to a concentration of 7.80 x 10^{12} GC/mL based on the declared titer. The plot with a detection wavelength of 260 nm can be found in Fig. 2A.

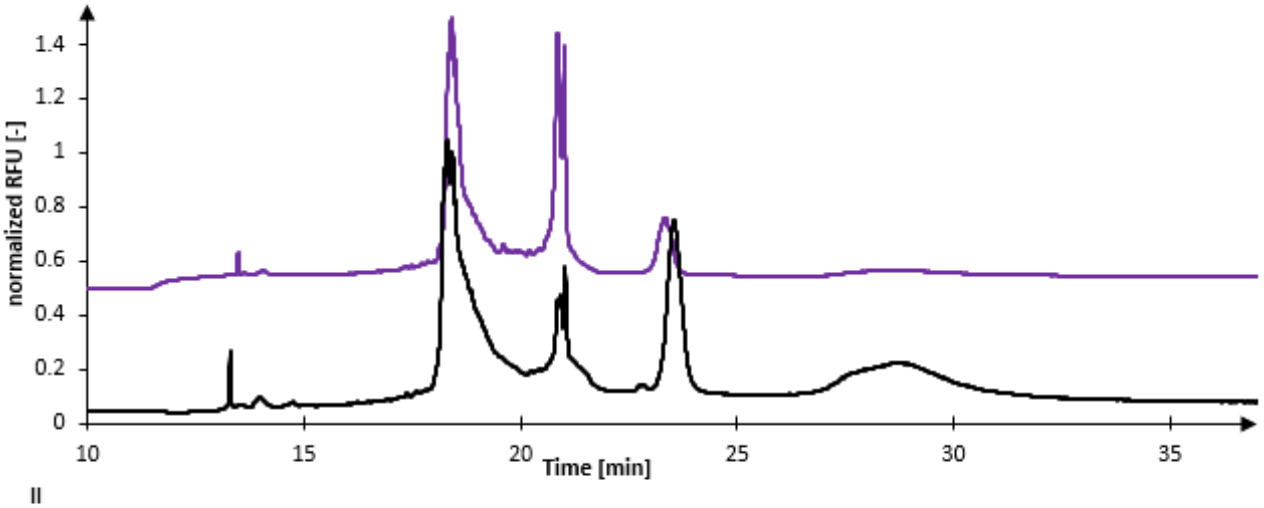


Fig. S2. Comparison of CGE-LIF separation of full AAV3 S2 purified with the QIAquick PCR purification kit and heated at 70 °C for 2 min, followed by 5 min in ice (black trace) and same sample but adding formamide (final concentration 10%) followed by the same conditions of 70 °C, 2 min and 5 min in ice (purple trace). RFU values were normalized for the highest value within one trace. For further conditions, see Figure 3.

2ND PROJECT:

2nd Publication

Reversed phase liquid chromatography for recombinant AAV genome integrity assessment

Analytical Chemistry. 2023 May; 95(22):8478-86

analytical chemistry

pubs.acs.org/ac

Article

Reversed Phase-Liquid Chromatography for Recombinant AAV Genome Integrity Assessment

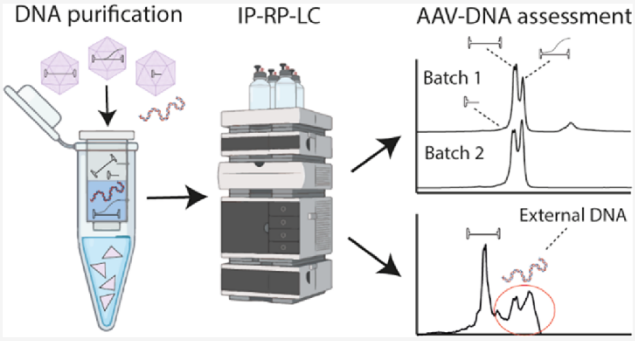
Christoph Gstöttner,[⊥] Andrei Hutanu,[⊥] Sacha Boon, Aurelia Raducanu, Klaus Richter, Markus Haindl, Raphael Ruppert,[#] and Elena Domínguez-Vega*[,#]

Cite This: https://doi.org/10.1021/acs.analchem.3c00222

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: After decades of research, gene therapy products have reached market maturity in recent years. Recombinant adeno-associated viruses (rAAVs) are one of the most promising gene delivery vehicles and are currently under intense scientific investigation. These next-generation medicines remain very challenging when it comes to designing appropriate analytical techniques for quality control. One critical quality attribute is the integrity of ssDNA incorporated in these vectors. The genome is the active compound driving rAAV therapy and therefore requires proper assessment and quality control. Current techniques for rAAV genome characterization include next-generation sequencing, quantitative polymerase chain reaction, analytical ultracentrifugation (AUC), and capillary gel electrophoresis (CGE), yet each of them presents their limitations or lack of user-friendliness. In this work, we demonstrate for the first time the potential of ion pairing-reverse phase-liquid chromatography (IP-RP-LC) to characterize the integrity of rAAV genomes. The obtained results were supported by two orthogonal techniques, AUC and CGE. IP-RP-LC can be performed above DNA melting temperatures, avoiding the detection of secondary DNA isoforms, and does not require the use of dyes due to UV detection. We demonstrate that this technique is suitable for batch comparability, different rAAV serotypes (AAV2 and AAV8), internal vs external (inside vs outside the capsid) DNA analysis, and contaminated samples. Overall, it is exceptionally user-friendly, needs limited sample preparation, has high reproducibility, and permits fractionation for further peak characterization. All of these factors add significant value of IP-RP-LC to the analytical toolbox of rAAV genome assessment.



INTRODUCTION

Gene therapy products using viral delivery vehicles, such as recombinant adenovirus, lentivirus, or adeno-associated virus (AAV), are recently gaining broad attention in the pharmaceutical industry. In particular, rAAVs combine several attributes that make them predestined for use as a gene delivery vehicle. From a safety perspective, the site-specific genome integration,[1] lack of pathogenicity in humans,[2] and dependency on helper viruses[3] are very favorable. Additionally, there are multiple rAAV serotypes with different tropisms for a specific tissue permitting it to be used to treat diseases very specifically.[4] Until 2022, already 136 clinical trials of recombinant AAV (rAAV) in different phases were reported, with two rAAV therapies for retinal dystrophy and spinal muscular atrophy authorized by the FDA.[5] AAVs are nonenveloped viruses containing a protein capsid built of three different proteins, namely, VP1, VP2, and VP3. The wild-type capsid contains a single-stranded DNA (ssDNA) genome with a length of up to 4.8 kilobase pairs (kbp).[6] This ssDNA usually contains two open reading frames, coding for proteins necessary for replication, capsid formation, and assembly.[7] Both ends of the ssDNA of AAVs are flanked by inverted terminal repeats (ITRs) that form dsDNA hairpin loops. They serve as primers for DNA polymerase and thus are crucial for genome replication. Furthermore, they are necessary for the loading of the genome into the capsid.[8] For therapeutic purposes, AAV vectors are recombinantly produced, and the aforementioned genes in between the two ITRs can be exchanged by a transgene cassette, preventing rAAV from replicating in patients.[7] After entering the target cells and delivering ssDNA into the nucleus, double-stranded DNA (dsDNA) episomes are formed.[9,10] These are stable and lead to a long-term expression of the gene of interest in patients allowing treatment of genetic diseases.[11−13]

Received: January 14, 2023
Accepted: May 3, 2023

In addition to the intended genome, rAAV preparations can contain several other DNA impurities, which can either be located outside the capsid and are copurified during production or inside the capsid. While the external DNA might potentially contain host cell DNA, it can also consist of plasmids used for rAAV production and, thereby, resistance cassettes that are necessary for cell selection during rAAV production.[14] The copurified DNA is, in most of the cases, removed by a Benzonase or DNase I treatment during downstream processing.[15,16] These species should be monitored carefully due to their potential genotoxicity.[17] Another source of external DNA can be encapsidated DNA, which can be ejected from the capsid when samples are stressed.[18,19] Next to the external, also internal DNA can contain selection markers but also fragments or truncated forms of the intended genome.[20−23] All of these species are potentially immunogenic and/or genotoxic for patients.[17] For this reason, profound characterization of the rAAV genome is of the utmost importance to ensure a safe and effective product. Another critical quality attribute (CQA) relates to the rAAV genome, which is transferred to human cells and therefore needs to be monitored carefully. Generally, only the region flanked by the two ITRs is loaded into the rAAV capsids; however, some reports describe the encapsidation of the host cell or plasmid backbone DNA.[17,21,23] This can be especially concerning since the production plasmids contain antibiotic resistance genes and other sequences, which may lead to the expression of immunogenic proteins or peptides.[24] These impurities range from 1 to 5%[21] but can account for the majority (>80%) of the genome when the unfavorable rep/cap helper plasmid and transfer plasmid combinations are used.[20] Some of them can be a product of reverse transcription of the plasmid necessary for rAAV production. This can be minimized by using plasmids that have backbones exceeding the maximum loading capacity of around 4.8 kbp of AAVs.[24]

rAAV-DNA impurities are currently characterized by quantitative real-time PCR (qPCR),[25,26] next-generation sequencing (NGS),[27,28] capillary gel electrophoresis (CGE),[18,29] and analytical ultracentrifugation (AUC).[18,20] qPCR is very sensitive and able to detect single DNA molecules. However, a disadvantage of qPCR is its biased approach relying on specific primers that can only detect certain DNA sequences and therefore is at risk of missing noncommonly observed DNA impurities.[30,31] NGS can give a good overview of the genome integrity, with single-molecule real-time (SMRT) sequencing being the gold standard for rAAVs due to its ability to sequence entire rAAV genomes without the need for in silico reconstruction.[27,28,32] The generation of dsDNA (necessary to ligate the SMRTbells) can be achieved either by annealing self-complementary rAAV-DNA or by extension of the 3′ end by terminal transferase. One drawback of this technique is the bias toward smaller molecules, therefore under-representing longer DNA fragments.[27] In addition, DNA fragments that are unable to ligate to a SMRTbell adapter can be missed.[27,28] For these reasons, complementary techniques able to directly analyze the ssDNA genome without any need for further sample manipulation are required. In contrast to the mentioned techniques, AUC analyzes intact unmodified viruses and allows the evaluation of empty and filled capsids based on their different densities and different sedimentation properties. Thus, it can be used to distinguish between full and empty capsids but also between rAAV capsids containing ssDNA with large size differences (2.1 vs 4.3 kbp).[20,33] However, a downside of AUC is that it requires high volumes of rAAV samples.[33] Overall, qPCR, NGS, and AUC require a high level of expertise but also long sample preparation and analysis time.[30] CGE has demonstrated to be able to separate different DNA species found in rAAV and thereby represents a fast and economic possibility to test the genome integrity.[18,29] Yet, no full denaturation of dsDNA to ssDNA can be achieved, so under commonly used analysis conditions, some dsDNA artifacts can form, complicating the data analysis.[34] In addition, CGE has the benefit of very low sample consumption with injection volumes in the nanoliter range. However, for low-concentrated samples, this can be a drawback, as it limits sensitivity, especially when UV detection is employed. Moreover, the low sample volume represents a significant hurdle when it comes to fractionation, limiting further characterization options. Ion pairing reversed-phase liquid chromatography (IP-RP-LC) has, due to its high user-friendliness and robustness, also great potential to characterize DNA samples.[35] The separation mechanism is based on the use of a lipophilic ion-pair reagent, such as triethylamine (TEA), which binds to the negatively charged DNA phosphate backbone. This ion-paring effect results in an increased hydrophobicity with an increase in DNA length, permitting the separation of differently-sized DNA species in an RP column. Another influence on the separation of different DNA species results from the different hydrophobicity of the DNA bases.[36] Up to date, IP-RP-LC has been applied to analyze RNA samples[37,38] as well as short and medium-sized DNA[35,39] (up to 100 bp) fragments. Here, we demonstrate for the first time the potential of IP-RP-LC to assess the genome integrity of rAAVs with transgene sizes ranging from 2.5 to 4.6 kbp and compare our results with the two orthogonal techniques, AUC and CGE.

EXPERIMENTAL SECTION

Samples and Chemicals. Reagents and materials used for this study were at least of analytical grade; for more details, see Supporting Information Method S1. AAV2-V, AAV8-V, and AAV8-HEK samples were purchased from supplier 1, while AAV2-S was from supplier 2.

Preparation of the Plasmid and rAAV-DNA for IP-RP-LC. 10 or 20 μL of the rAAV sample (titers between 8×10^{12} and 2×10^{13} vg) was mixed with 50 μL of PB buffer (QIAquick PCR purification kit, Qiagen), loaded onto a QIAquick column, and centrifuged at 16,100g for 1 min. Afterward, the sample was washed with 750 μL of PE buffer (QIAquick PCR purification kit, Qiagen) and centrifuged for 1 min. After discarding the flow-through and centrifugation for an additional 1 min, the sample was eluted with 50 μL of elution buffer (QIAquick PCR purification kit, Qiagen). Before analysis, the sample was heat treated at 95 °C for 2 min, followed by 5 min incubation on ice. For DNA digestion, Benzonase was diluted 1:10 in a 10× digestion buffer (100 mM Tris−HCl, 25 mM $MgCl_2$, 5 mM $CaCl_2$ at a pH 7.6). 17 μL of the purified DNA sample was mixed with 2 μL of 10× digestion buffer and 1 μL of the diluted Benzonase. The mixture was incubated for 1 or 2 h, and digestion was stopped by heating the samples to 95 °C for 2 min, followed by 5 min cool-down on ice. For plasmid digestion, 5 μg of plasmid DNA was mixed with the provided digestion buffer as instructed by the manufacturer. 10 U of the enzyme per μg of DNA in a total volume of 100 μL was used and incubated for 60 min at 37 °C.

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

The absence of proteins was confirmed using RP-LC (Supporting Information Method S2).

Analysis of the Plasmid and rAAV-DNA with IP-RP-LC. For the analysis of plasmid or rAAV-DNA, an Agilent 1200 series instrument equipped with a quaternary pump (G1311A) combined with a degasser (G1322A), an autosampler (G1367D) with a thermostat (G1330B), a column oven (1316B), and a variable wavelength detector (G1314C) with a standard cell (Agilent Technologies, Waldbronn, Germany) was employed. Analysis of rAAV-DNA or the plasmid digest was performed using a DNA-PAC column at 95 °C with 0.1 M TEAA in water pH 7 (adjusted with TEA) as mobile phase A and 0.1 M TEAA in 75% H_2O with 25% ACN pH 7 as mobile phase B. The starting condition was 35% B, which was gradually increased to 70% B in 18 min, followed by an increase to 100% B in 2 min. After 5 min of cleaning of the column at 100% B, the starting conditions were reached during a 3 min gradient, followed by re-equilibration of the column for 7 min, resulting in a total analysis time of 35 min with a constant flow rate of 0.4 mL/min. The injection volume was either 20 μL per purified ssDNA rAAV sample or 3 μL digested plasmid DNA. rAAV-ssDNA was detected using UV detection at 260 nm. The runs were aligned using a FastRuler Middle Range DNA Ladder, which was measured together at the beginning and end of each sequence.

Preparation and Analysis of the Plasmid and rAAV-DNA for CGE-LIF. 5 or 20 μL of the rAAV sample was prediluted with 1× phosphate-buffered saline (PBS) (Sigma-Aldrich) as required for the specific analysis. Afterward, samples (rAAV and digested plasmid) were purified following the instructions from the QIAquick PCR purification kit but with two washing steps of the QIAquick column. Before injection in CGE, the sample was heated at 70 °C for 2 min, followed by 5 min on ice. For analysis, a SCIEX PA800 Plus system (Brea) equipped with a solid-state laser with an excitation wavelength of 488 nm and a 520 nm bandpass emission filter (Cat. no. 65-699) from Edmund Optics (Barrington) with a temperature-controlled autosampler (±2 °C) was used. Data were acquired with 32 Karat software 10.3. Separation was performed in a bare fused silica capillary with a 50 μm internal diameter and 20 cm effective and 30 cm total length. As gel buffer 2% PVP, 4 M urea in a 1× TBE solution with 1:25,000 diluted SYBR green II was used.[29,40,41] Voltage was set at −6 kV. The electrokinetic injection was performed for 30 or 60 s at a voltage of −5 kV. The capillary temperature was set to 25 and 10 °C was used for the autosampler.

Preparation and Analysis of rAAV Samples Using Sedimentation Velocity Analytical Ultracentrifugation (SV-AUC). Before analysis, the samples were thawed and equilibrated at room temperature, followed by the measurement of UV absorbance spectra (using a NanoDrop One UV−Vis spectrophotometer from Thermo Fisher Scientific) in the wavelength range of 220−350 nm to confirm a suitable initial sample concentration. Afterward, samples were diluted to an optical density (at 230 nm) of 0.8 except for AAV8 produced in HEK cells, which already had an optical density below 0.8. For measurements of sedimentation velocity, an Optima analytical ultracentrifuge from Beckman-Coulter (Brea, California) with an 8-hole AN-50 Ti analytical rotor and 12 mm charcoal epon double-sector centerpieces was used. Experimental conditions are detailed in Supporting Information Method S3.

RESULTS AND DISCUSSION

Development of the IP-RP-LC Method for rAAV Genome Characterization. The genomic material in AAV-based gene therapy products is encapsidated within three structural proteins, namely, VP1, VP2, and VP3. As these proteins can influence the separation performance in IP-RP-LC by coeluting with some ssDNA species, purification of DNA from the rAAV sample is required prior to analysis. For this purpose, we performed a silica membrane-based DNA purification step using commercially available spin columns. For the approach, we use a limited amount of the rAAV sample for purification (10 or 20 μL with a titer of 2×10^{13} to 8×10^{12} vg, respectively), which resulted in 50 μL of purified DNA. This allowed duplicate analysis from a single DNA purification with 20 μL injection volume for each measurement. To ensure complete removal of the protein, the sample was analyzed before and after purification using RP-LC (Figure S1). Before DNA purification, the proteins could be detected, while no proteins were visible after DNA purification, excluding any interference during the DNA separation.

Genomes of rAAVs are ssDNA; however, due to the nature of the hairpin-shaped ITRs, the DNA molecule can have either sense or antisense polarity.[42] When the DNA of rAAVs is purified for analysis, dsDNA, as well as sophisticated multiplexes, may form alongside the ssDNA genome. Therefore, it is important to avoid all forms of dsDNA as these sample preparation artifacts can add complexity to the analysis. To reduce additional signals arising from these species, the use of high temperatures following a fast cooling on ice is often used.[18] In our case, we heated the sample to 95 °C (above the melting temperature of dsDNA, around 85−90 °C depending on the sequence) for 2 min, followed by 5 min incubation on ice prior to analysis. The high temperatures should melt dsDNA into ssDNA and then lead to the formation of internal double strands in the ITR region of the rAAV genomes due to the rapid cooling and potentially limiting annealing between sense and antisense strands. However, as shown by Hutanu et al.,[18] this approach is not fully efficient, and still, some dimers can form, resulting in additional signals. In IP-RP-LC, oligonucleotides and RNA are commonly analyzed at a temperature of 50 °C,[38,43,44] with few reports showing that with an increased temperature up to 80 °C, resolution can be improved.[37,39] Huang et al. showed that increasing the temperature from 50 to 65 °C simplified the number of RNA conformation structures detected.[44] To avoid rearrangements and ensure the detection of rAAV and its impurities in the single-stranded form, we evaluate the possibility of performing the analysis above DNA melting temperatures. Figure 1 shows a comparison of an AAV8-V DNA sample analysis at 55, 75, and 95 °C. At 55 °C, a main signal between 17 and 19 min was observed. Because this temperature favors the formation of dsDNA, this signal most likely represents dsDNA species. Increasing the temperature to 75 °C resulted in two clusters of signals at 12 and 17 min, suggesting the coexistence of two populations during the separation (ssDNA and dsDNA, respectively). Since 75 °C is below the melting temperature, only a small fraction of DNA is in the single-stranded form. This is also supported by the separation mechanism of IP-RP: dsDNA pairs with a higher number of TEA molecules, resulting in increased hydrophobicity, and therefore, later elution compared to ssDNA. Increasing the temperature above the melting point (95 °C) resulted in only

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

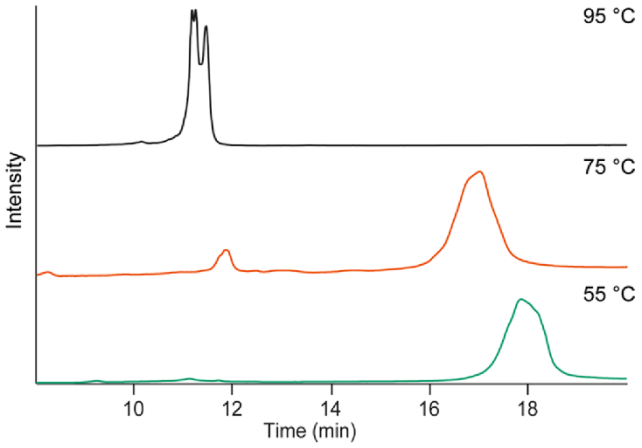


Figure 1. IP-RP-LC analysis of the purified AAV8-V genome at 55 °C (green trace), 75 °C (orange trace), and 95 °C (black trace).

one cluster of signals around 11−12 min, which reflects multiple ssDNA species. Similarly to the profile observed at 75 °C, the large shift in elution time observed (from 18 to 11.5 min) corresponded to the change of the ssDNA conformation (lower hydrophobicity) rather than to the increase in temperature itself, which only had a minor influence. Additionally, the peak shape and resolution improved at higher column temperatures, as previously shown in the literature for small DNA fragment analysis.[39] As a consequence, the broad peak observed at 55 °C resolved in multiple signals at 95 °C corresponding to different ssDNA species present in the rAAV sample.

To further confirm that these signals arise from the genomic material, we analyzed the corresponding AAV8-V empty capsid sample (containing no DNA) from the same batch and performed an identical sample preparation. Figure S2 shows a comparison between AAV8-V full and empty samples using an analysis temperature of 95 °C. The AAV8-V empty sample showed only minor signals (zoom Figure S2), indicating that the detected peaks in the AAV8-V full sample come from the loaded genome. To get more information on the resolving power of the method, a sample consisting of fragments of different sizes was analyzed. To this end, a plasmid with a length of 5118 bp was digested using *Apa*LI, which resulted in four fragments with sizes of 2070 bp, 1305 bp, 1246 bp, and 497 bp. IP-RP-LC analysis showed four distinct signals separated by the length of DNA due to the ion pairing with TEA, resulting in a higher hydrophobicity the longer the DNA fragments are (Figure 2A). Under the applied conditions, a baseline separation of the two fragments, which differed only in 59 base pairs in size, could be achieved. To corroborate our results, we analyzed the same plasmid digest using a CGE-LIF approach published recently.[18,29] CGE-LIF showed very similar profiles compared to IP-RP-LC with more efficient peak shapes for CGE-LIF, which is an intrinsic characteristic of this technique. In IP-RP-LC analysis, the peak at 497 bp showed a double peak. This peak was also broader in the CGE electropherogram (Figure 2B), also indicating a possible mixture of two DNA fragments such as a clipped variant or an unspecific cleavage. Additionally, while for the CGE-LIF method, an SYBR green dye, which binds ssDNA, is necessary, our IP-RP-LC method can directly detect DNA by UV. This opens the possibility that DNA can be measured at 95 °C, which is not possible with SYBR green due to the missing

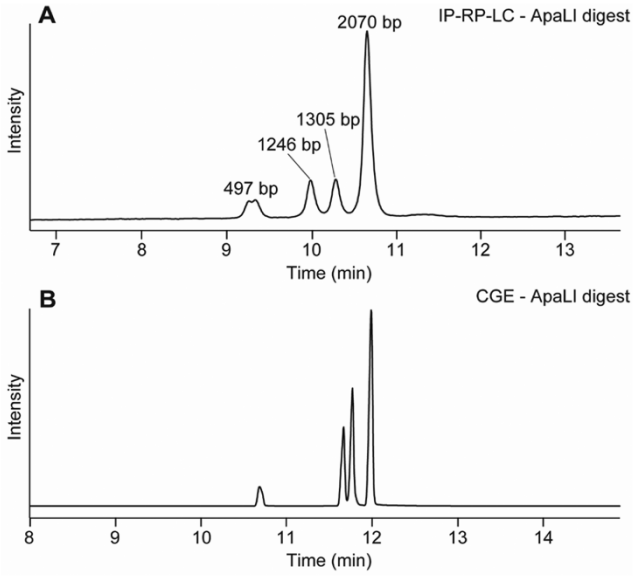


Figure 2. Analysis of the *Apa*LI plasmid digest by (A) IP-RP-LC and (B) CGE-LIF.

interaction between DNA and dye at these temperatures. Furthermore, LC has some benefits in the view of user-friendliness and reproducibility[36] and opens the possibility of peak fractionation for further characterization (which may be challenging by CGE-LIF due to the low volumes employed). We also analyzed a digest of the same plasmid with *Sna*BI and *Xba*I, resulting in two fragments of 366 and 4752 bp (Figure S3). The 366 bp fragment eluted at around 9 min, while 4752 bp eluted much later (around 11.5 min) as expected. Overall, CGE and IP-RP-LC analysis resulted in very comparable results with higher peak efficiency for CGE. Finally, we analyzed one of the most complex AAV samples (AAV2-V) over the course of 3 days, resulting in a very similar profile and retention times for all three measurements on the 3 days with RSD values ≤ 4.7% for the AAV genome peak (11.2 min) (Figure S4).

Discrimination between Internal and External DNA Using IP-RP-LC in Combination with Benzonase Treatment. During downstream processing, AAV samples are often treated with Benzonase in order to remove the potential DNA material present outside the capsid.[45] Yet, in some cases, the remaining external DNA material can be observed in AAV preparations. Therefore, we evaluated if we could discriminate between internal and external DNA by combining our method with a Benzonase treatment prior to analysis.

As a test sample, we selected a rAAV, which showed a very complex IP-RP-LC profile with multiple signals, which were suspected to be external DNA (Figure 3). This sample was a rAAV from serotype 2 (AAV2-V), which should contain a genome of around 2.5 kb. After analysis, the expected genome was detected at around 11 min together with several additional larger DNA species (Figure 3, black trace). To investigate the location of these DNA impurities (external or internal), two Benzonase digestions were performed. One before and one after the disassembly of the capsids and subsequent DNA purification. After incubation for 2 h with Benzonase before capsid disassembly (Figure 3 green trace), a decrease in the intensity of larger-sized species was observed, in particular for the peaks between 12 and 14 min. These results suggest that

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

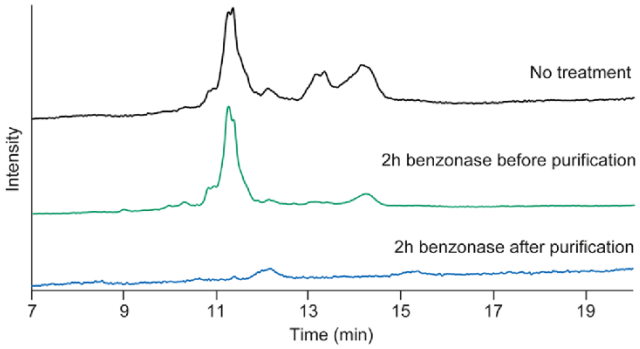


Figure 3. IP-RP-LC analysis of the AAV2-V genome at 95 °C without any Benzonase treatment (black trace), after 2 h Benzonase treatment before DNA purification (green line), and 2 h Benzonase treatment after DNA purification (blue trace).

these DNA species are indeed not packed into the AAV2-V particles and therefore digested, whereas the rAAV genome is resistant to Benzonase digestion due to its protection by the protein capsid. For the peak eluting at around 14.5 min, a significant decrease of signal was observed, yet a minor signal remained present. Due to its late elution, this signal should correspond to a DNA species with a size above 10 kb, which cannot be integrated into AAV. Therefore, we speculate that the remaining signal may arise from incomplete digestion or from other non-DNA impurities. AAV2-V was also incubated with Benzonase for 2 h after capsid disassembly and DNA purification (Figure 3 blue trace). After this procedure, no signal for the rAAV genome was observed. Only a minor signal, around 12 min, which was resistant to Benzonase digestion, was detected in the chromatogram. This signal was also observed in the AAV2-V empty sample. For this reason, we concluded that these peaks do not correspond to DNA but rather to minor protein impurities.

IP-RP-LC for Integrity Assessment of Different-Sized AAV Genomes. AAVs can bear diverse genomes, and their genomic impurities can vary from sample to sample. Therefore, we purchased rAAV samples loaded with different-sized genomes and degrees of heterogeneity and applied the developed IP-RP-LC approach for their characterization. During method development, we observed an extraordinary complexity of rAAV samples, often leading to ambiguous results and difficult data interpretation. To increase comprehensiveness, we decided to flank our LC technique with CGE and SV-AUC. Since all techniques use very different separation principles, a more accurate picture of the genome integrity can be retrieved when combining them.

AAVs can accommodate up to 4.8 kbp of the genomic material. To evaluate the performance of the IP-RP-LC approach, we analyzed two AAV8 loaded with different length ssDNA. One of the selected AAV8 samples was expressed in HEK293 cells and was loaded with a 4.6 kbp genome. Figure 4A shows the IP-RP-LC separation obtained for AAV8-HEK with one main peak at 12 min representing the 4.6 kb ssDNA genome and some shorter DNA fragments eluting upfront. CGE-LIF showed a very similar profile with a higher number of smaller fragments separated from the main peak. In particular larger fragments migrating around 12.5 min were not resolved with IP-RP-LC (Figure 4B). The SV-AUC analysis perfectly matches the results obtained by the other two approaches, with one main peak at 105 S, as well as some

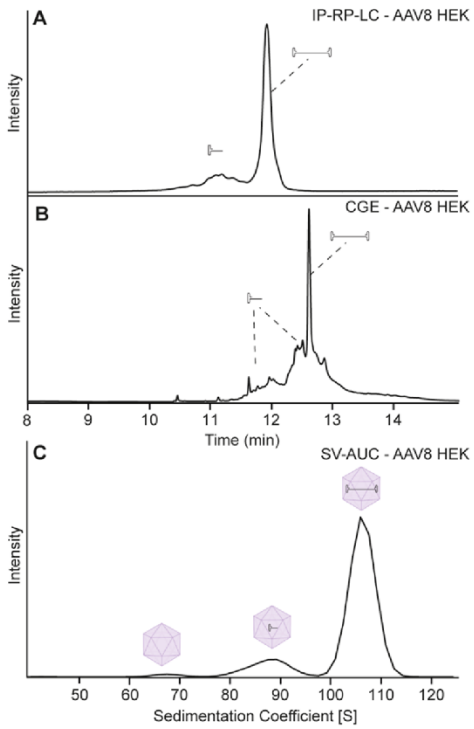


Figure 4. Analysis of AAV8-HEK after DNA purification by (A) IP-RP-LC and (B) CGE-LIF or (C) analysis of the intact rAAV particles with SV-AUC.

particles with shorter DNA at 80−95 S (Figure 4C). In contrast to IP-RP-LC and CGE, SV-AUC is also capable of visualizing empty capsids at around 65 S.

For a shorter genome, we selected a rAAV8-V material produced in Sf9 insect cells and containing a theoretically loaded genome of approximately 2.5 kbp (AAV8-V Batch 1). Analysis of the AAV8-V batch 1 sample with IP-RP-LC showed a more complex profile comprising two distinct peaks, which were partially separated, as well as small signals eluting before and after the main peak (Figure 5A). A similar picture could also be observed with CGE-LIF, where two high abundance peaks were partially separated (Figure 5B). In IP-RP-LC, additionally, a partial resolution of the first high abundance peak was observed. We speculate that most probably they correlate to two ssDNA fragments with a very similar length but different hydrophobicity (e.g., sense and antisense), which were not resolved with CGE-LIF or SV-AUC, as their separation mechanisms are not affected by the hydrophobicity of the molecule. The species with faster mobility corresponds to the 2.5 kbp target DNA and the slower signal to a higher-sized species. SV-AUC further confirmed the findings as it also showed a very clear separation of two distinct species. Capsids containing 2.5 kbp DNA were detected around 90 S, and the heavier species were observed at 105 S (Figure 5C). The longer ssDNA construct most likely represents an elongation of the ssDNA genome with part of the plasmid backbone to the maximum loading capacity of a rAAV particle. Similar findings were also observed by another study using charge detection mass spectrometry as well as mass photometry.[46] This effect might be due to the fact that shorter ssDNA has not had the optimal loading length for rAAV particles and therefore can result in unwanted elongation with the production plasmid backbone. Because this could lead to the transfer of some

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

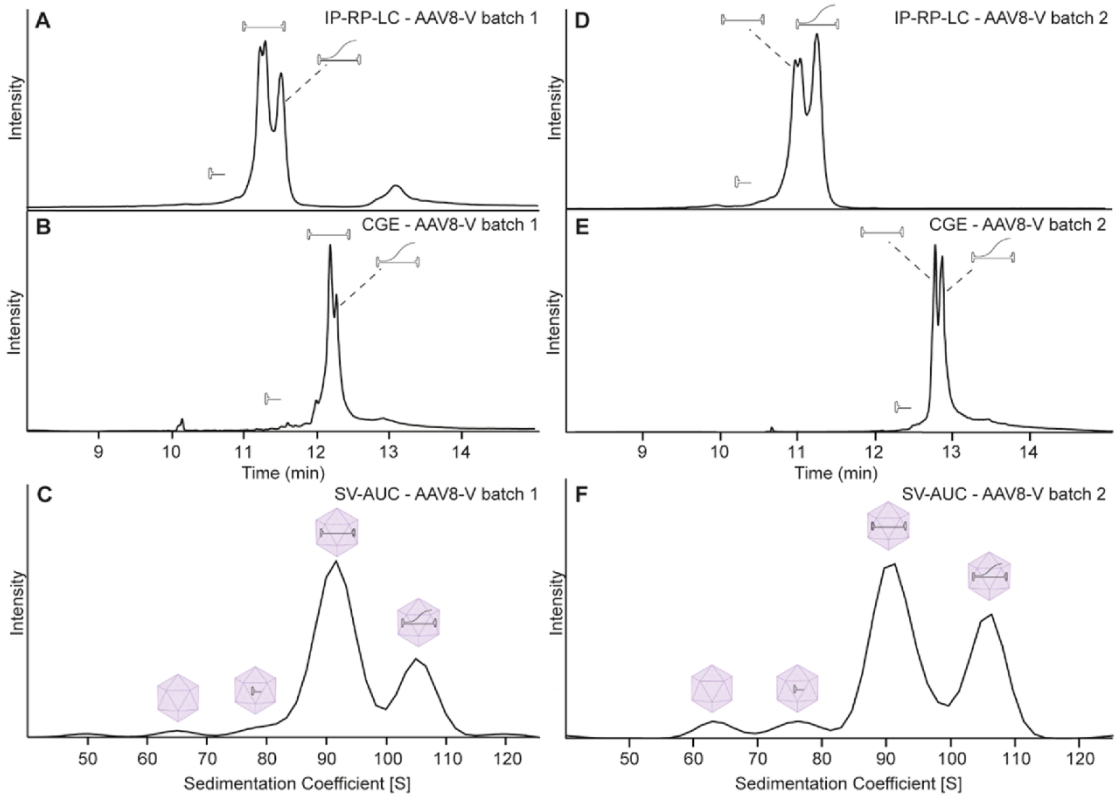


Figure 5. Analysis of AAV8-V batch 1 after DNA purification by (A) IP-RP-LC and (B) CGE-LIF or of the intact rAAV particles with (C) SV-AUC, as well as AAV8-V batch 2 by (D) IP-RP-LC, (E) CGE-LIF, and (F) SV-AUC.

selection marker genes into the patient, these species should be very closely monitored and kept at a minimum to guarantee a safe rAAV product.

Next to the elongated genome, we detected some additional low abundance signals before the main peak corresponding to shorter DNA fragments. These low abundance signals are more clearly detectable by CGE-LIF. Also, in SV-AUC, capsids containing smaller DNA fragments were detected at 75−80 S. AAVs containing shorter DNA fragments are often observed in rAAV samples and have been previously reported.[28] These rAAV capsids incorporate only a part of the expected genome, and they may lead to lower efficiency of the gene therapy product due to the lack of the target genome and can potentially be immunogenic and/or genotoxic for the patient, therefore requiring close monitoring. Finally, following the main peak, some signals were detected with CGE-LIF and IP-RP-LC. Due to the large size of these signals, we suspected that they correspond to external DNA, which was not completely removed by the Benzonase treatment during downstream processing, similar to the case of AAV2-V (Figure 3). Also, the fact that these signals are not detected by AUC suggests that they are not packed into the capsid; otherwise, a signal larger than 110 S would be expected.

In view of all of these observations, we also analyzed a second batch of AAV8 loaded with the 2.5 kbp genome (AAV8-V Batch 2). The IP-RP-LC results showed a similar picture; however, it seems that the larger DNA signal at 13 min is not present in this batch anymore. This again suggests that this large DNA was indeed external DNA from the production process, which was not properly removed in batch 1 (Figure 5D). Similarly, CGE-LIF also shows a slightly lower signal for the larger DNA species (Figure 5E), with the remaining ones most likely being dsDNA artifacts as previously reported by Hutanu et al.[18] While the LC data suggest a lower amount of the elongated genome in batch 1, batch 2 shows nearly a 1:1 ratio. This finding was supported by CGE-LIF data as well as SV-AUC data (Figure 4F). Overall, these results demonstrate that the IP-RP-LC method is very well suited for batch comparability studies and thus adds an additional technique to the analytical toolbox of rAAVs.

As a last case, we applied the IP-RP-LC method to detect rAAV genomic contaminations. The AAV-2S sample (3.3 kbp) contained contamination by another AAV2 strain with a different genome (4.2 kbp, as indicated upfront by the supplier). The IP-RP-LC profile of the AAV-2S sample showed two clear main peaks and an additional small signal around 10.5 min (Figure 6A). The first peak represents the intended genome with a length of around 3.3 kbp, whereas the second represents the contaminant genome with a length of around 4.2 kbp. Furthermore, fragmented DNA was detected before the main peak, which was higher in intensity compared to the previously analyzed AAV8 samples. CGE-LIF largely supports these results by detecting two peaks, as well as small fragments, before the main peak (Figure 6B). Next to these two main peaks, an additional peak in the CGE profile at 12 min was observed, which was not detected by IP-RP-LC or SV-AUC. It cannot be excluded that this is a dsDNA or multiplex artifact due to incomplete denaturation or as a consequence of the very low concentration of this sample. Additional artifacts were observed throughout the complete analysis in CGE-LIF (Figure S5 marked with *), whereas in IP-RP-LC, no comparable signals were detected (only an increase in the baseline between 23 and 30 min due to the gradient). To exclude carryover effects as a cause for the unexpected signals,

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

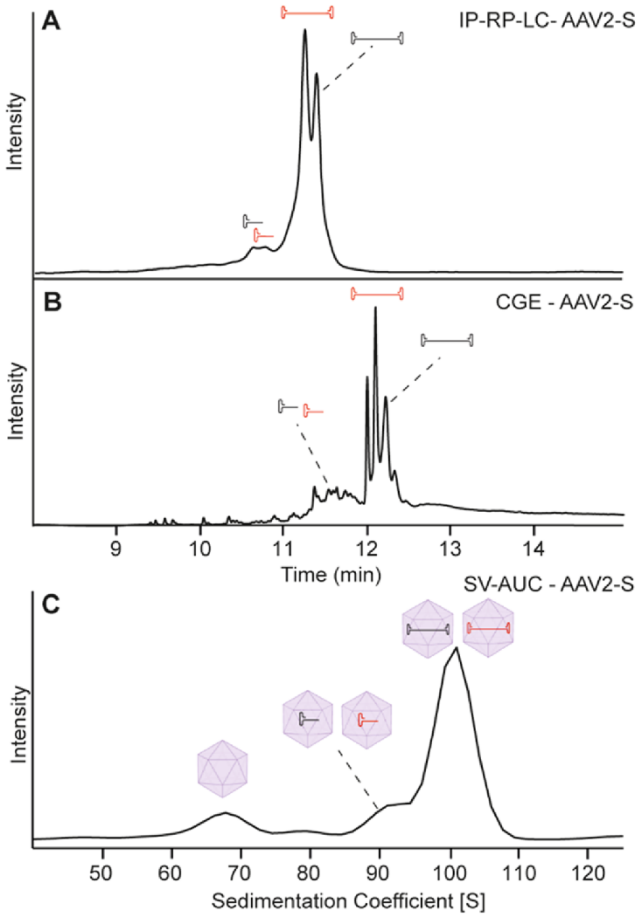


Figure 6. Analysis of AAV2-S after DNA purification by (A) IP-RP-LC and (B) CGE-LIF or of the intact rAAV particles with (C) SV-AUC.

blank injections were performed before and after each CGE analysis. These injections were blank for two different analyses on two different devices, so we concluded that the cause for the observed signals must lie within the specimen (not shown). In numerous rAAV samples not mentioned in this study, we observed that CGE tends to have more artifacts in complex samples, in general. It should be mentioned that for this AAV2-S sample, the absolute intensity was slightly lower when compared to other rAAV analyses (most likely due to the lower virus titer), which could also play a role in the discrepancy between LC and CGE for this sample. Looking at the SV-AUC data, empty capsids were observed between 65 and 70 S, followed by some rAAV particles containing shorter DNA fragments between 75 and 95 S (Figure 6C). In contrast to IP-RP-LC and CGE-LIF, only one peak at 100 S was detected. These results illustrate the strength of SV-AUC to discriminate between different species and give a complete picture of the capsid filling state. However, smaller differences in genome length (only around 800 bp) may be challenging to detect by SV-AUC using standard conditions, illustrating once again the complementarity of these techniques.

CONCLUSIONS

We have developed an IP-RP-LC approach for the analysis of the genomic material loaded in rAAV samples. We demonstrated the capacity of the approach to characterize the denatured ssDNA material from rAAV samples in a simple, fast, economic, and user-friendly way. In order to identify strengths and limitations, we compared our approach with the established techniques CGE-LIF[18,29] and SV-AUC[18,20] in a variety of rAAV samples. Results between all techniques were mostly consistent, which highlights the general applicability of the proposed method. Running the analysis at 95 °C allows to melt dsDNA and thereby analyze only ssDNA in contrast to the CGE-LIF approach, which also detects and separates dsDNA[18] and other secondary structures.[29] By using UV detection, we avoid the drawbacks of DNA dyes and have a direct detection of target DNA. Furthermore, LC is already established in the pharmaceutical industry for biopharmaceutical and small molecule drug characterization and therefore can be easily implemented. Comparison with CGE-LIF and SV-AUC supported our findings on the presence of elongated genomes and fragments of the genome of interest in a variety of rAAV samples. When the material is scarce, SV-AUC is not a viable option (approx. 400 μL consumption for a low viral titer), while IP-RP-LC requires only around 10−20 μL per injection, and CGE can perform multiple injections out of the same amount. IP-RP-LC was able to separate the genome of a rAAV2-S sample from a contaminant differing in 800 bp. Here, the IP-RP-LC and CGE-LIF approaches showed their benefit over the SV-AUC method, which was not able to resolve smaller differences in ssDNA under standard conditions. Overall, CGE-LIF provided higher peak efficiency compared to IP-RP-LC. On the other hand, IP-RP-LC allows peak fractionation and further characterization (e.g., by NGS) in a straightforward way. Furthermore, with some adaptations (e.g., the use of volatile ion pairing reagents or integration in 2D-LC with an MS-compatible second dimension), IP-RP-LC could be hyphenated with mass spectrometry for direct peak identification. Overall, the proposed IP-RP-LC is a very powerful alternative for the genome integrity assessment of rAAVs that can be easily adopted in QC labs, thus complementing the analytical toolbox of these next-generation therapeutics.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c00222.

RP-LC analysis of VP proteins before and after purification; comparison of AAV8-V full and empty samples by IP-RP-LC; analysis of the plasmid digest by IP-RP-LC and CGE-LIF; analysis of AAV2-V in triplicates; IP-RP-LC chromatogram of AAV-2V and blank and comparison to CGE-LIF; and additional experimental information (PDF)

(PDF)

AUTHOR INFORMATION

Corresponding Author

Elena Domínguez-Vega − *Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333ZA Leiden, The Netherlands;* orcid.org/0000-0002-6394-0783; Email: e.dominguez_vega@lumc.nl

Authors

Christoph Gstöttner − *Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333ZA Leiden, The Netherlands*

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

Andrei Hutanu − *Pharma Technical Development Analytics, F. Hoffmann-La Roche AG, 4070 Basel, Switzerland;* orcid.org/0000-0003-0174-5250
Sacha Boon − *Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333ZA Leiden, The Netherlands*
Aurelia Raducanu − *Pharma Technical Operation Cell- and Gene Therapy, Roche Diagnostics GmbH, 82377 Penzberg, Germany*
Klaus Richter − *Coriolis Pharma Research GmbH, 82152 Planegg, Germany*
Markus Haindl − *Pharma Technical Operation Cell- and Gene Therapy, Roche Diagnostics GmbH, 82377 Penzberg, Germany*
Raphael Ruppert − *Pharma Technical Operation Cell- and Gene Therapy, Roche Diagnostics GmbH, 82377 Penzberg, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c00222

Author Contributions

[⊥]C.G. and A.H. are shared first authors.

Author Contributions

[#]R.R. and E.D.-V. are shared last authors.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Henckaerts, E.; Dutheil, N.; Zeltner, N.; Kattman, S.; Kohlbrenner, E.; Ward, P.; Clément, N.; Rebollo, P.; Kennedy, M.; Keller, G. M.; Linden, R. M. *Proc. Natl. Acad. Sci.* **2009**, *106*, 7571−7576.
(2) Monahan, P. E.; Jooss, K.; Sands, M. S. *Expert Opin. Drug Saf.* **2002**, *1*, 79−91.
(3) Meier, A. F.; Fraefel, C.; Seyffert, M. *Viruses* **2020**, *12*, No. 662.
(4) Vandenberghe, L. H.; Wilson, J. M.; Gao, G. *Gene Ther.* **2009**, *16*, 311−319.
(5) Au, H. K. E.; Isalan, M.; Mielcarek, M. *Front. Med.* **2022**, *8*, No. 2746.
(6) Naso, M. F.; Tomkowicz, B.; Perry, W. L., 3rd; Strohl, W. R. *BioDrugs* **2017**, *31*, 317−334.
(7) Ayuso, E.; Mingozzi, F.; Bosch, F. *Curr. Gene Ther.* **2010**, *10*, 423−436.
(8) Earley, L. F.; Conatser, L. M.; Lue, V. M.; Dobbins, A. L.; Li, C.; Hirsch, M. L.; Samulski, R. J. *Hum. Gene Ther.* **2020**, *31*, 151−162.
(9) Choi, V. W.; McCarty, D. M.; Samulski, R. J. *J. Virol.* **2006**, *80*, 10346−10356.
(10) Daya, S.; Berns, K. I. *Clin. Microbiol. Rev.* **2008**, *21*, 583−93.
(11) Penaud-Budloo, M.; Le Guiner, C.; Nowrouzi, A.; Toromanoff, A.; Chérel, Y.; Chenuaud, P.; Schmidt, M.; von Kalle, C.; Rolling, F.; Moullier, P.; Snyder, R. O. *J. Virol.* **2008**, *82*, 7875−85.
(12) Bijlani, S.; Pang, K. M.; Sivanandam, V.; Singh, A.; Chatterjee, S. *Front. Genome Ed.* **2022**, *3*, No. 42.
(13) Bulcha, J. T.; Wang, Y.; Ma, H.; Tai, P. W. L.; Gao, G. *Signal Transduction Targeted Ther.* **2021**, *6*, No. 53.
(14) Schnödt, M.; Schmeer, M.; Kracher, B.; Krüsemann, C.; Espinosa, L. E.; Grünert, A.; Fuchsluger, T.; Rischmüller, A.; Schleef, M.; Büning, H. *Mol. Ther.–Nucleic Acids* **2016**, *5*, No. e355.
(15) Kimura, T.; Ferran, B.; Tsukahara, Y.; Shang, Q.; Desai, S.; Fedoce, A.; Pimentel, D. R.; Luptak, I.; Adachi, T.; Ido, Y.; Matsui, R.; Bachschmid, M. M. *Sci. Rep.* **2019**, *9*, No. 13601.
(16) Guan, J.-S.; Chen, K.; Si, Y.; Kim, T.; Zhou, Z.; Kim, S.; Zhou, L.; Liu, X. *Front. Chem. Eng.* **2022**, *4*, No. 1.
(17) Wright, J. F. *Biomedicines* **2014**, *2*, 80−97.
(18) Hutanu, A.; Boelsterli, D.; Schmidli, C.; Montealegre, C.; Dang Thai, M. H. N.; Bobaly, B.; Koch, M.; Schwarz, M. A. *Electrophoresis* **2022**, *43*, 1107−1117.
(19) Xu, Y.; Guo, P.; Zhang, J.; Chrzanowski, M.; Chew, H.; Firrman, J. A.; Sang, N.; Diao, Y.; Xiao, W. *Mol. Ther.–Methods Clin. Dev.* **2020**, *18*, 328−334.
(20) Burnham, B.; Nass, S.; Kong, E.; Mattingly, M.; Woodcock, D.; Song, A.; Wadsworth, S.; Cheng, S. H.; Scaria, A.; O'Riordan, C. R. *Hum. Gene Ther: Methods.* **2015**, *26*, 228−242.
(21) Recombinant Adeno-Associated Virus. *Mol. Ther.* **2003**, *7*, S348−S349.
(22) Hauck, B.; Murphy, S. L.; Smith, P. H.; Qu, G.; Liu, X.; Zelenaia, O.; Mingozzi, F.; Sommer, J. M.; High, K. A.; Wright, J. F. *Mol. Ther.* **2009**, *17*, 144−152.
(23) Chadeuf, G.; Ciron, C.; Moullier, P.; Salvetti, A. *Mol. Ther.* **2005**, *12*, 744−753.
(24) Wright, J. F. *Gene Ther.* **2008**, *15*, 840−848.
(25) D'Costa, S.; Blouin, V.; Broucque, F.; Penaud-Budloo, M.; Fçranois, A.; Perez, I. C.; Le Bec, C.; Moullier, P.; Snyder, R. O.; Ayuso, E. *Mol. Ther.–Methods Clin. Dev.* **2016**, *3*, No. 16019.
(26) Wang, Y.; Menon, N.; Shen, S.; Feschenko, M.; Bergelson, S. *Mol. Ther.–Methods Clin. Dev.* **2020**, *19*, 341−346.
(27) Tai, P. W. L.; Xie, J.; Fong, K.; Seetin, M.; Heiner, C.; Su, Q.; Weiand, M.; Wilmot, D.; Zapp, M. L.; Gao, G. *Mol. Ther.–Methods Clin. Dev.* **2018**, *9*, 130−141.
(28) Tran, N. T.; Heiner, C.; Weber, K.; Weiand, M.; Wilmot, D.; Xie, J.; Wang, D.; Brown, A.; Manokaran, S.; Su, Q.; Zapp, M. L.; Gao, G.; Tai, P. W. L. *Mol. Ther.–Methods Clin. Dev.* **2020**, *18*, 639−651.
(29) Luo, J.; Guttman, A. *LCGC Suppl.* **2020**, *33*, 33−38.
(30) Lecomte, E.; Tournaire, B.; Cogné, B.; Dupont, J. B.; Lindenbaum, P.; Martin-Fontaine, M.; Broucque, F.; Robin, C.; Hebben, M.; Merten, O. W.; Blouin, V.; François, A.; Redon, R.; Moullier, P.; Léger, A. *Mol. Ther.–Nucleic Acids* **2015**, *4*, No. e260.
(31) Lecomte, E.; Tournaire, B.; Cogné, B.; Dupont, J.-B.; Lindenbaum, P.; Martin-Fontaine, M.; Broucque, F.; Robin, C.; Hebben, M.; Merten, O.-W.; Blouin, V.; François, A.; Redon, R.; Moullier, P.; Léger, A. *Mol. Ther.–Nucleic Acids* **2015**, *4*, No. e260.
(32) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; deWinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. *Science* **2009**, *323*, 133−138.
(33) Maruno, T.; Usami, K.; Ishii, K.; Torisu, T.; Uchiyama, S. *J. Pharm. Sci.* **2021**, *110*, 3375−3384.
(34) McLaren, R. S.; Ensenberger, M. G.; Budowle, B.; Rabbach, D.; Fulmer, P. M.; Sprecher, C. J.; Bessetti, J.; Sundquist, T. M.; Storts, D. R. *Forensic Sci. Int.: Genet.* **2008**, *2*, 257−273.
(35) Yin, J.; Zhang, N.; Wang, H. *TrAC, Trends Anal. Chem.* **2019**, *120*, No. 115645.
(36) Santos, I. C.; Brodbelt, J. S. *J. Sep. Sci.* **2021**, *44*, 340−372.
(37) Azarani, A.; Hecker, K. H. *Nucleic Acids Res.* **2001**, *29*, No. e7.
(38) Close, E. D.; Nwokeoji, A. O.; Milton, D.; Cook, K.; Hindocha, D. M.; Hook, E. C.; Wood, H.; Dickman, M. J. *J. Chromatogr. A* **2016**, *1440*, 135−144.
(39) Oefner, P. J. *J. Chromatogr. B: Biomed. Sci. Appl.* **2000**, *739*, 345−355.
(40) Chang, H.-S.; Wanders, B.; Guttman, A. *BioTechniques* **2002**, *32*, 1228−1230.
(41) Hutanu, A.; Boelsterli, D.; Schmidli, C.; Montealegre, C.; Dang Thai, M. H. N.; Bobaly, B.; Koch, M.; Schwarz, M. A. *Electrophoresis* **2022**, *43*, 1107−1117.
(42) Mayor, H. D.; Torikai, K.; Melnick, J. L.; Mandel, M. *Science* **1969**, *166*, 1280−1282.

(43) Huber, C. G.; Oefner, P. J.; Bonn, G. K. *Anal. Biochem.* **1993**, *212*, 351−358.
(44) Huang, Z.; Jayaseelan, S.; Hebert, J.; Seo, H.; Niu, L. *Anal. Biochem.* **2013**, *435*, 35−43.
(45) Wolf, M. W.; Reichl, U. *Expert Rev. Vaccines* **2011**, *10*, 1451−75.
(46) Ebberink, E. H. T. M.; Ruisinger, A.; Nuebel, M.; Thomann, M.; Heck, A. J. R. *Mol. Ther.–Methods Clin. Dev.* **2022**, *27*, 491−501.

https://doi.org/10.1021/acs.analchem.3c00222
Anal. Chem. XXXX, XXX, XXX−XXX

SUPPORTING INFORMATION

Reverse Phase liquid chromatography for recombinant AAV genome integrity assessment

Analytical Chemistry. 2023 May; 95(22):8478-86

Supporting information Figures

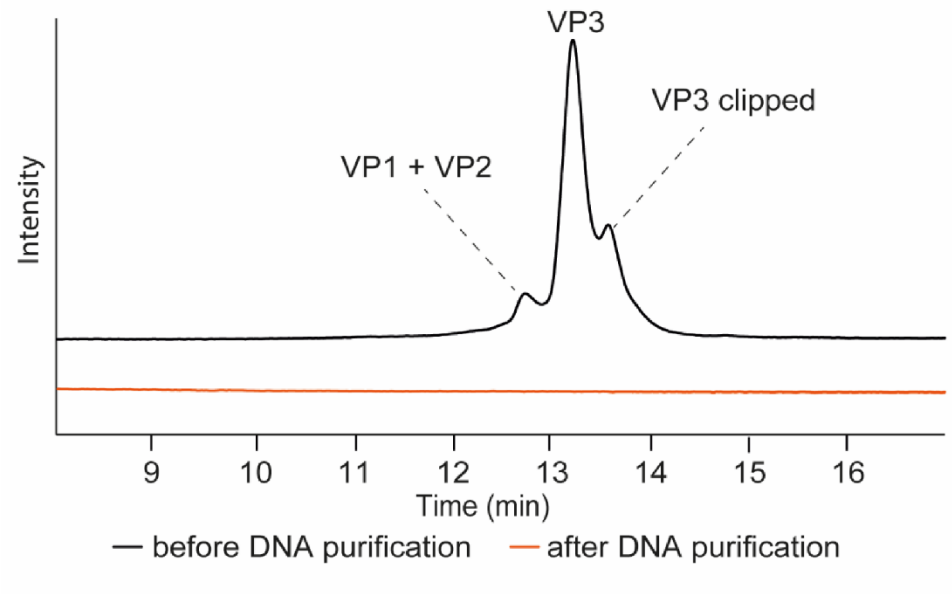


Figure S1: RP-LC analysis of a AAV8-V full sample before DNA purification (black line) and after DNA purification (orange line).

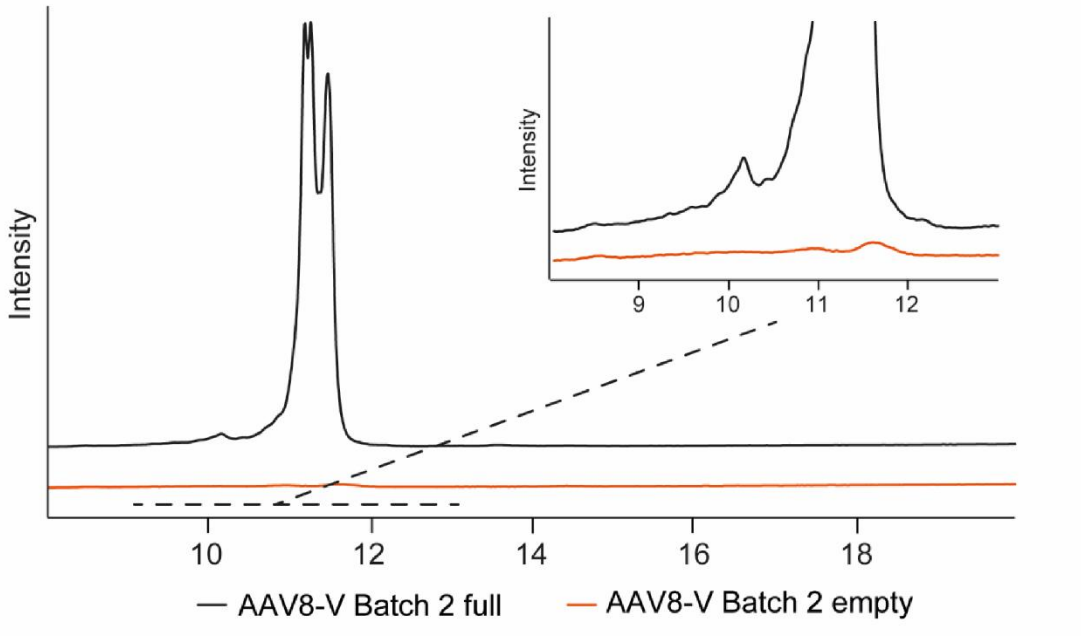


Figure S2: IP-RP-LC analysis of a AAV8-V full sample (black line) and a AAV8-V empty sample orange line. Additionally, a zoom in the elution region is depicted.

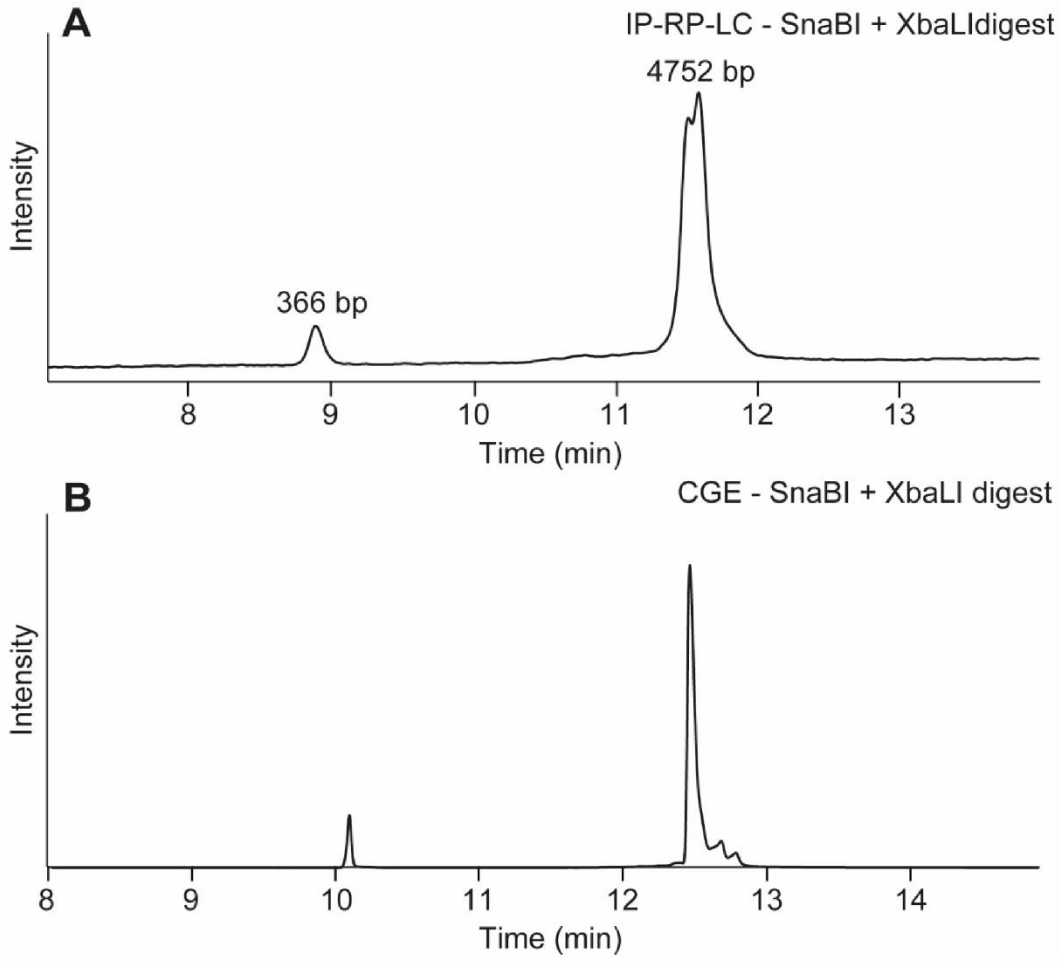


Figure S3: Analysis of a SnaBI + XbaLI plasmid digest by A) IP-RP-LC and B) CGE-LIF.

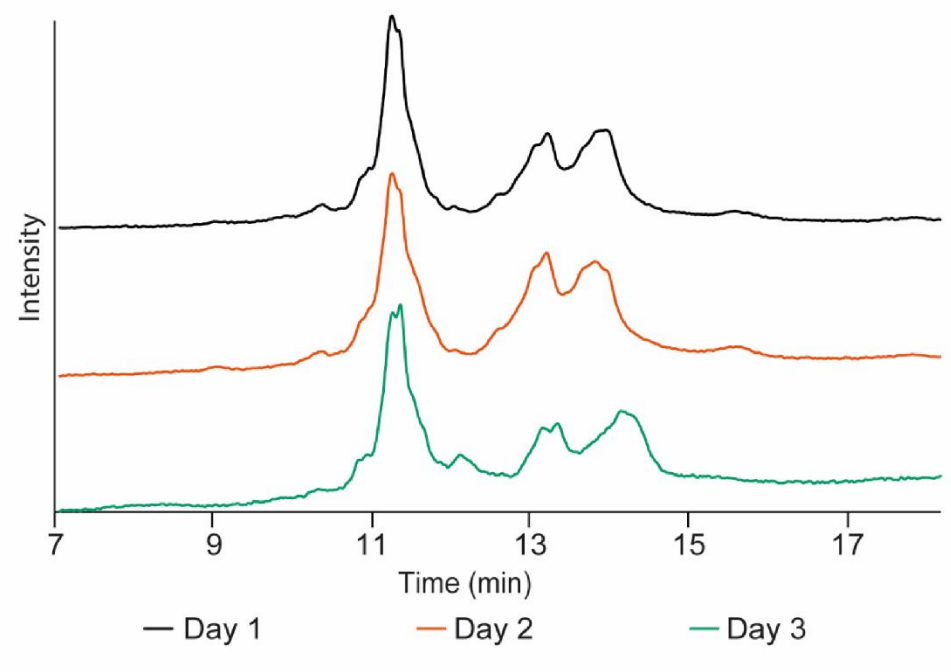


Figure S4: IP-RP-LC analysis of AAV2-V on three different days.

Supporting information methods

Information S1 – Reagents and materials

Acetonitrile was provided by Actuall Chemicals (Oss, The Netherlands). For all mobile phases, demineralized water from an Elga water purification system was used. 1 M aqueous solution of triethylammonium acetate (TEAA) was purchased from Fluka (Buchs, Switzerland). The used SYBR™ Green II RNA Gel Stain 10,000 x concentrate in DMSO, UltraPure™ DNase/RNase-Free Distilled Water, FastRuler Middle Range DNA Ladder and the DNAPac RP column 2.1 x 100 mm with a particle size of 4 µm were obtained from Thermo Fisher Scientific (Waltham, Massachusetts). The BioResolve RP mAb Polyphenyl column 2.1 x 150 mm with a particle size of 2.7 µm was purchased from Waters (Milford, Massachusetts). The QIAquick PCR purification kit was purchased from Qiagen (Hilden, Germany). Benzonase Nuclease, $MgCl_2$ and $CaCl_2$ dihydrate were from Sigma Aldrich (St. Louis, Missouri). Phosphate buffered saline (PBS) tablets (1 tablet in 1L water results in 140 mM NaCl, 10 mM phosphate buffer, 3 mM KCl, pH 7.4), glacial acetic acid, Polyvinylpyrrolidone (PVP), 10x tris(hydroxymethyl)aminomethane (Tris) borate EDTA (TBE) buffer, Tween 20, difluoroacetic acid (DFA), Hellmanex 3 and Urea were supplied by Sigma-Aldrich/Merck KGaA (Darmstadt; Germany). Tris-HCl was obtained from Invitrogen (Waltham, Massachusetts). Bare fused silica CE capillaries with an inner diameter of 50 µm were purchased from Polymicro Technologies /Molex LLC (Phoenix, USA). Plasmid DNA (Cat. no. VB190926-1395dab) was provided by VectorBuilder Inc. (Chicago, USA). ApaLI, SnaBI and XbaI restriction endonucleases were obtained from New England Biolabs (Ipswich, USA). 10x PBS (1.55 M NaCl, 27 mM sodium phosphate dibasic, 15 mM KCl, pH 7.2) and 10% Pluronic F68 were obtained from Gibco Thermo Fisher Scientific (Waltham, Massachusetts). Ethanol absolute was purchased from VWR (Darmstadt, Germany). Fill hole gaskets (neoprene) and fill hole screws were provided by Spin Analytical (South Berwick, Maine).

Information S2 - Analysis of AAV proteins with RP-LC

For the analysis of rAAV proteins the AAV sample was either disassembled using 10% acetic acid in the case of direct analysis or purified with the QIAquick PCR purification kit and thereby disassembled in the using the provided binding buffer (Buffer PB). For the analysis an Agilent 1200 series instrument equipped with a quaternary pump (G1311A) combined with a degasser (G1322A), autosampler (G1367D) with thermostat (G1330B), column oven (1316B) and variable wavelength detector (G1314C) with a standard cell was used (Agilent Technologies, Waldbronn, Germany) was employed. Analysis rAAV proteins either before or after purification was performed using a BioResolve RP mAb Polyphenyl column 2.1 x 150 mm column at 80 °C with 0.1% DFA in water as mobile phase A and 0.1% DFA in ACN as mobile phase B. The starting condition was 35% B which was gradually increased to 50% B in 25 min followed by an increase to 95% B in 1 min. After 4 min cleaning of the column at 95% B the starting conditions were reached during a 1 min gradient followed by a re-equilibration of the column for 9 min resulting in a total analysis time of 40 min with a constant flowrate of 0.15 mL/min. The injection volume was 10 µL. The proteins were detected using UV detection at 280 nm.

Information S3 - Analysis of rAAV by sedimentation-velocity analytical ultracentrifugation (SV-AUC)

For measurements of sedimentation velocity an Optima analytical ultracentrifuge from Beckman-Coulter (Brea, California) with an 8-hole AN-50 Ti analytical rotor and 12-mm charcoal-epon double-

sector centerpieces was used. Cleaning upfront AUC measurements was performed using AUC cell washers (Spin Analytical, South Berwick, Maine), and was confirmed by two consecutive absorbance intensity scans through empty cells. Afterwards the sample sector was filled with 380 µL rAAV sample and the reference sector with 400 µL formulation buffer. The formulation buffer consisted of 1 x PBS (Gibco) and 0.001% Pluronic F-68 in case of AAV 8 samples or 1 x PBS (Gibco) with 0.014% Tween 20 in case of AAV2 samples. After a temperature equilibration of sample and rotor with resting rotor, the sedimentation profiles were recorded using the absorbance detection optics in intensity mode (at 230 nm and 280 nm) at 12000 rpm. Scans 1-50 were analyzed using UltraScan III using a partial specific volume of 0.73 mL/g and a radial data range from approximately 0.03 cm above meniscus position to 7.10 cm. The raw data were pre-processed (meniscus position fit, time invariant and radial invariant noise corrections) by using 2-dimensional spectrum analysis (2DSA). Afterwards, the obtained results were refined by using parametrically constrained spectrum analysis (PCSA) applying the straight line (SL) model. Estimation of confidence intervals was performed by 100 Monte Carlo (MC) iterations on top of the PCSA-SL results. The calculations were carried out at Coriolis Pharma Research GmbH on an in-house Boston Server cluster with 128-core AMD Epyc™ 7702 processors. Pseudo-3D distributions were analyzed with a sample-specific s-value range bin set to determine the content of empty capsids, full capsids, light molecular weight species (LMWS) and heavy molecular weight species (HMWS).

2.3 Using Peptide Nucleic Acid Hybridization Probes for Qualitative and Quantitative Analysis of Nucleic Acid Therapeutics by Capillary Electrophoresis

The analysis of rAAV vectors revealed that methods for NA are not always straight forward and that the industry is currently seeking new approaches that complement the analytical repertoire. A lot of fundamental work on CE for NAs has been done in the 1990s but remained rather unnoticed. Throughout my research the idea came up to use fluorescent oligonucleotides as probes in CE. A detailed literature research revealed an article from O'Keefe and colleagues [185] that initiated working analytically with peptide nucleic acids (PNAs). PNAs are synthetic organic polymers that possess chemical properties that are characteristic of DNA and RNA. The primary distinguishing feature is the presence of a backbone composed of N-(2-aminoethyl)glycine monomers, which are connected by peptide bonds. [186] The biological stability of PNAs is attributed to their uncharged nature, which renders them resistant to digestion by proteases and nucleases. The absence of electrostatic repulsion provides a greater attraction towards complementary NA sequences, compared to normal NAs. This enables hybridization conditions that are outside the reach of conventional nucleic acid oligomers, such as low levels of sodium and high temperatures, among others. The electrophoretic separation of PNA and hybridized NA is feasible due to the significant difference in their charge-to-hydrodynamic radius ratio. [187] Thus, PNA oligomers were utilized as probes in CE to identify particular DNA sequences. By labeling the PNAs with fluorescent labels, a detailed and sensitive monitoring of hybridization was possible. The investigation involved a series of proof-of-concept trials that aimed at determining the capacity of PNA probes to detect NAs in new therapeutic entities, both qualitatively and quantitatively. The method obtained was suitable for a broad range of NA sizes, making it valuable for drug product formulations with varying levels of intricacy, including oligonucleotides, mRNA vaccines, or rAAVs. The method was highly reproducible and determined the quantity and identity of ssNAs within a broad range of 20 to 1000 nt. The concept was successfully implemented for mRNA, with a LOQ in the picomolar range achieved through the use of multiple probes. However, in ds samples only fragments identical to PNA size could be measured. This constraint could be overcome by means of enzymatic digestion and by employing multiple probes.

3RD PROJECT:

Using Peptide Nucleic Acid Hybridization Probes for Qualitative and Quantitative Analysis of Nucleic Acid Therapeutics by Capillary Electrophoresis

Analytical Chemistry. 2023 Mar; 95(11):4914-22..

analytical chemistry

pubs.acs.org/ac Article

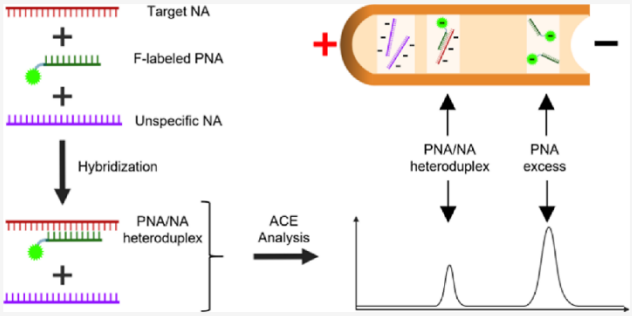
Using Peptide Nucleic Acid Hybridization Probes for Qualitative and Quantitative Analysis of Nucleic Acid Therapeutics by Capillary Electrophoresis

Andrei Hutanu,* Chiara Signori, Bernd Moritz, Manuel Gregoritza, Adelheid Rohde, and Maria A. Schwarz

Cite This: *Anal. Chem.* 2023, 95, 4914−4922 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The space of advanced therapeutic modalities is currently evolving in rapid pace necessitating continuous improvement of analytical quality control methods. In order to evaluate the identity of nucleic acid species in gene therapy products, we propose a capillary electrophoresis-based gel free hybridization assay in which fluorescently labeled peptide nucleic acids (PNAs) are applied as affinity probes. PNAs are engineered organic polymers that share the base pairing properties with DNA and RNA but have an uncharged peptide backbone. In the present study, we conduct various proof-of-concept studies to identify the potential of PNA probes for advanced analytical characterization of novel therapeutic modalities like oligonucleotides, plasmids, mRNA, and DNA released by recombinant adeno-associated virus. For single-stranded nucleic acids up to 1000 nucleotides, the method is an excellent choice that proved to be highly specific by detecting DNA traces in complex samples, while having a limit of quantification in the picomolar range when multiple probes are used. For double-stranded samples, only fragments that are similar in size to the probe could be quantified. This limitation can be circumvented when target DNA is digested and multiple probes are used opening an alternative to quantitative PCR.

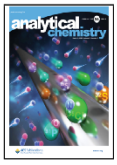


INTRODUCTION

There is an estimate of 5000−10,000 rare diseases for the human population, and a significant part of them is caused by genetic mutations.[1,2] After decades of research, health authorities approved the first gene therapy products for monogenic diseases like spinal muscular atrophy or Leber's congenital amaurosis,[3−5] and there are efforts to use vectors such as the recombinant adeno-associated virus (rAAV) as a generic platform to tackle a wide range of rare diseases.[6] However, the application of gene delivery medications is still extremely challenging, as there are several obstacles that need to be overcome. From a clinical perspective, immunogenicity, efficacy, genotoxicity, and genomic persistence are key factors that need to be considered when designing gene therapies.[7−10] Ensuring high product quality is essential to address safety concerns linked to immunogenicity and genotoxicity and is, thus, currently a field under intense investigation. Several reviews have suggested critical quality attributes that need to be addressed to safeguard top-class gene therapy products.[11−14]

From an analytical perspective, these drug products tend to be heterogeneous and highly complex since there is both a delivery vehicle (virus and lipids) and a nucleic acid (NA) payload (DNA/RNA) required for most applications. When dealing with single-stranded NAs (ssNAs) that have more than 200 nucleotides (nt), options to confirm their identity and quantity are very limited. This is due to the large size itself (> kDa range), the high variability of instable secondary structures, and the increased enzymatic susceptibility in the case of RNA. Physicochemical techniques such as hydrophilic interaction liquid chromatography, anion-exchange chromatography, or ion-pairing reversed phase liquid are powerful in the low size range,[15] but evaluation of identity by elution time of long NA fragments often remains challenging.[16] The situation is similar for capillary electrophoresis (CE)-based methods such as capillary gel electrophoresis. Even though the separation efficiency is greater, analysis of ssNAs show difficulties with regard to robust and reproducible separations.[17] Therefore, polymerase chain reaction (PCR)-based techniques are typically used for NA identity and quantifica-

Received: October 31, 2022
Accepted: February 2, 2023
Published: March 8, 2023



tion.[18−20] Their great benefit is a very high sensitivity and the general abundance of the necessary equipment. On the other side, handling can be challenging as cross-contamination with amplifiable DNA is possible.[20] In practice, just a small piece of the target NA is actually identified, which might cause misleading results when only one set of primers is used. This can be avoided by next-generation sequencing (NGS) where all NAs in an analyte are completely identified.[21,22] In this way, a holistic view of the state of a sample is gathered. However, NGS remains complex as extensive sample preparation and data analysis are required in comparison to PCR. As a result, it is mainly used in early development and not for routine testing. In the case of RNA, both methods need a reverse transcription step, which can add a hidden bias. The rapidly rising number of gene therapy candidate drugs, their wide variety,[23] and increasingly accelerated development times underscore the need for novel, fast, versatile, and direct analytical approaches.

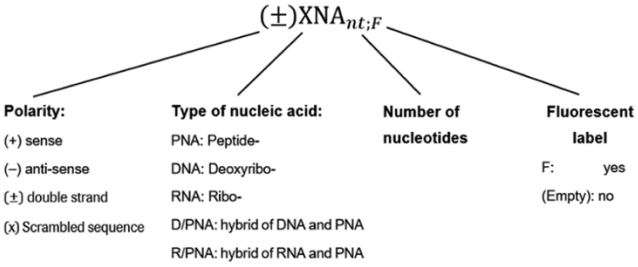
The goal of our work was to develop a fast, accurate, and inexpensive method for identification and quantitation of NAs of various origins. As a methodological platform, we chose a combination of target−probe hybridization with separation of hybrids from the unreacted probes by gel-free CE. Hereby, peptide nucleic acids (PNAs) were chosen as probes. This approach has been previously used by Hu et al.[24] for the determination of multiple miRNAs in the low mass range (<25 nt). We propose a method in which PNAs are applied as hybridization probes binding to a wide size range of NAs in the presence of complementary or nonspecific NA strands. PNAs are engineered organic polymers that share chemical properties with DNA and RNA. The main difference is the backbone that consists of N-(2-aminoethyl)glycine monomers, which are linked through peptide bonds while the secondary amine group is substituted with an acetate derivative of a nucleobase.[25] The uncharged nature of PNAs makes them more biologically stable since proteases and nucleases cannot digest them. The missing electrostatic repulsion additionally gives them a higher affinity for complementary NA sequences than analogue NA oligomers.[26] This allows hybridization conditions that are not achievable with standard NA oligomers, for example, low sodium concentrations, high temperature, so forth. An electrophoretic separation between PNA and hybridized NA is easily realized since both macromolecules have a very different charge-to-hydrodynamic radius ratio. Fluorescently labeled PNA oligomers could, thus, be used as probes in CE for the detection of specific DNA sequences[24,27−32] as an alternative to PCR. In the present study, we conduct various proof-of-concept experiments to identify the potential of PNA probes for qualitative and quantitative detection of NAs in novel therapeutic modalities. To this end, optimal conditions for hybridization and separation in CE were screened and then applied to NA species that are currently of particular interest for the development of gene therapies, for example, oligonucleotides, plasmids, mRNA, and rAAVs.

EXPERIMENTAL SECTION

Reagents. Eurofins Genomics (Ebersberg; Germany) synthesized all DNA oligonucleotides. All PNAs were from PNABio (Newbury Park, USA), and the EGFP mRNA (cat. no. L-7601) was sourced by TriLink Biotechnologies (San Diego; USA). rAAV production plasmid DNA (cat. no. VB190926-1395dab) and a corresponding AAV2 sample were from VectorBuilder Inc. (Chicago, USA). For all the remaining reagents and the sequences of all nucleic acids used in this study (Scheme S1 and Tables S1 and S2), please refer to the Supporting Information.

Hybridization. Stock solutions were 25 μM in formamide for PNAs or in TE buffer (10 mM Tris; 1 mM EDTA) for DNA oligonucleotides. Screening for hybridization conditions was conducted in 15 mM Tris for a stable pH while various agents were tested (Table S3). The final hybridization conditions were 15 mM Tris and 2 mM sodium dodecyl sulfate (SDS) together with PNA, DNA, or mRNA in varying concentrations in a total volume of 30 μL. Up to 5% of formamide was accepted in the solution when high PNA excess was used. This mixture was heated to 95 °C for 10 min in a Thermomixer R (Eppendorf SE; Hamburg; Germany) and then cooled down slowly for 90 min until the block reached a temperature below 40 °C. For clarity, all the used nucleic acids were given a nomenclature as is described in Scheme 1. For details on the digest of plasmid and rAAV DNA, please refer to the Supporting Information.

Scheme 1. Nomenclature of Nucleic Acids Used in This Study



Capillary Electrophoresis. Analysis was carried out using a SciEx PA800 Plus system (Brea; USA) equipped with a solid-state laser with an excitation wavelength of 488 nm and a 520 nm band pass emission filter (cat. no. 65-699) from Edmund Optics (Barrington; USA), a 30 kV power supply, and a temperature-controlled autosampler (±2 °C). Data were acquired and analyzed using 32 Karat software 10.3. For the background electrolyte (BGE) optimization, 1× TBE was the base, and several agents were varied (Table S4). The composition of 1× TBE is 89.0 mM Tris, 89.0 mM boric acid, and 2.0 mM EDTA. The final BGE used in all PNA experiments after the screening was 6.6× TBE (587 mM Tris, 587 mM boric acid, and 13 mM EDTA), 3 M urea, and pH was controlled to be 8.3 ± 0.1. In some cases, 1:10,000 SYBR Green II was added to visualize unlabeled DNA as mentioned in the figure caption. For analysis, an eCAP neutral capillary, which uses a polyacrylamide coating (cat. no. 477441) from SciEx (Brea; USA) with an I.D. of 50 μm and cut to a total length of 30 and 20 cm length to window, was employed. The capillary cartridge was kept at 25 °C while samples were stored at 10 °C. Prior to each injection, the capillary was rinsed with water and 8 M urea and equilibrated with separation buffer at 345 kPa (50.0 psi) for 1 min each. The samples were injected at 6.9 kPa (1.0 psi) for 10 s, and the separation voltage was set to −30 kV reverse polarity. After each sequence, the capillary was rinsed with water for 10 min at 345 kPa. Data analysis of velocity corrected peak areas (CPAs): error bars represent the standard deviation from three measurements. CPAs were normalized by setting the highest value in a measurement as reference to 1.

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

RESULTS AND DISCUSSION

Basic Principles of Interaction between PNA and DNA Oligonucleotides. In our previous report,[33] we used orthogonal analytical techniques including transmission electron microscopy, analytical ultra centrifugation, and CGE to get a better idea of DNA heterogeneity inside rAAV capsids. The complex DNA purity profiles that were observed with CGE naturally caused the question to identify the unknown species. An article from Perry-ÒKeefe and colleagues showed an alternative to PCR or southern blotting by the use of fluorescently labeled PNAs as affinity probes. In this work, PNA-hybridized DNA (ss as well as dsDNA) of differently digested plasmids was used for qualitative analysis with CE, agarose gel electrophoresis, and CGE.[29] The basic principle of the technique is that in certain conditions, the PNA-to-DNA interaction (see Figure S1) is stronger than DNA-to-DNA bonding. In this case, a PNA sequence that is binding to one strand of a target DNA can be mixed and heated until DNA denatures. In the subsequent reannealing process, a PNA/DNA hybrid is formed, which can be detected due to fluorescent labeling of the PNA. The study has demonstrated that it is possible to select BGEs that keep PNA/DNA hybrids intact while DNA/DNA binding is inhibited, allowing detection of specific sequences in the sample. These results encouraged us to extend this technique on various forms of NAs. However, reports that actually used this PNA hybridization are scarce,[24,27−32] and we thus decided to begin method development from scratch to get a better understanding for crucial parameters.

Interaction of PNA and ssDNA. Our development process started in the simplest possible environment: our hypothesis was that a fluorescently labeled DNA 20mer ($(+)DNA_{20;F}$) should hybridize with an antisense and labeled PNA 15mer ($(-)PNA_{15;F}$) enabling subsequent analysis by CE. For hybridization, PNA and DNA oligonucleotides were mixed 1:1. The mixture was analyzed in a 1× TBE, 3 M urea BGE together with single injections of $(-)PNA_{15;F}$ and $(+)DNA_{20;F}$ (Figure S2A, upper three lines). However, the profile of the PNA was not reproducible, varying from hardly detectable (as shown in Figure S2A, upper red line) to clear peaks for different repetitions. DNA has a strongly negatively charged backbone that explains its fast migration. On the other hand, PNA is completely uncharged but carries a carboxyfluorescein (FAM) label that brings in a negatively charged carboxyl group, leading to migration in the electric field. Out of these results, we concluded that a mobility shift of DNA happened due to the formation of the DNA/PNA heteroduplex but that the PNA probe was not sufficiently soluble in the chosen conditions. For this reason, various agents were screened, including chaotropic salts, ionic and non-ionic detergents, and organic solvents that are known to increase the solubility of lipophilic molecules and are, in addition, compatible with CE. A detailed description regarding the effects of the tested agents can be found in Table S3. The screening showed most favorable results when adding 2 mM SDS to the hybridization reaction (Figure S2A, lower three lines). Under these conditions, the PNA profile was always reproducible, while DNA was not affected. Also, the hybrid peak shape was sharper and more stable. It is well reported in the literature that DNA hybrids are stabilized in sodium containing buffers and SDS is an agent that can be used for this purpose.[34] For the use of PNA as a probe, the dodecyl sulfate seemed to be particularly important for increasing the solubility of the PNA and thus accessibility for hybridization. The hybridization of PNA to DNA led to a minor migration time shift of the DNA. While the overall charge of the hybrid was probably similar to the one of ssDNA, the heteroduplex had a higher hydrodynamic radius. This migration time shift is expected to become smaller with increasing size of target DNA. To improve the resolution of ssDNA to hybrid and to get a better understanding of hybrid stability during electrophoresis, we decided to vary the BGE (see Table S4). Increasing salt concentration from 1× TBE to 6.6× TBE (Figure 1, upper

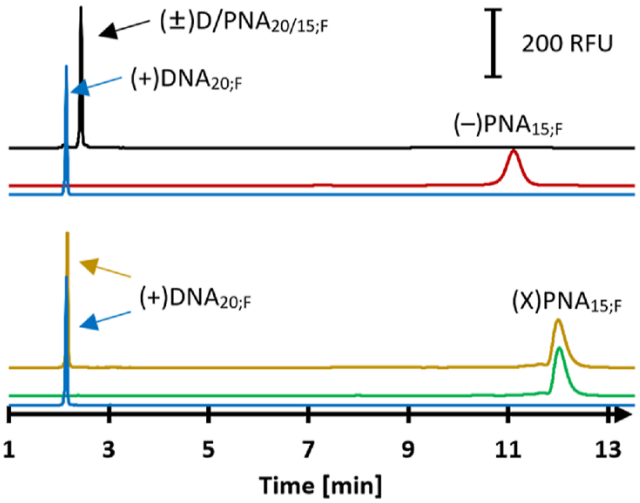


Figure 1. Hybridization of DNA and PNA oligonucleotides analyzed by CE. PNA and DNA signals are overlaid, while mixed samples are shifted on the *y*-axis for clarity. PNA and DNA samples were injected individually (upper blue and red line) or as a mixture (upper black line) at a concentration of 1 μM each. Hybridization takes place, and a shift of the DNA signal can be detected when incubated with PNA. Best separation results, in terms of signal intensity and resolution, were obtained with a BGE consisting out of 6.6× TBE with 3 M urea. The lower three lines serve as negative control: incubation of DNA with a PNA that has an unspecific sequence [$(X)PNA_{15;F}$] does not lead to a typical migration shift, and areas of PNA and DNA peaks are not affected.

three lines), a significant improvement in terms of signal intensity and resolution of heteroduplex from DNA was achieved. Through the increased resolution (Figure S2B, inset), it was possible to test the specificity of the approach by using a random PNA ($(X)PNA_{15;F}$) as a negative control. This PNA was mixed with the DNA sample. The mixture and the single components were also injected separately. No formation of hybrid was detected. The mixture of the two species led to an electropherogram that looked like the overlay of the single species (Figure 1B, lower three lines). After proving that the general concept works, we decided to apply the method to unlabeled DNA oligonucleotides. To evaluate the effect of DNA length, size was varied from 20 to 150 nt. The first goal was to detect the migration time of the unlabeled DNA by using a modified BGE, which contained SYBR Green II. Thereafter, the idea was to compare this with a run that did not use dye but labeled PNA to detect the hybrid. For this purpose, a two-fold DNA excess was used. Therefore, free DNA should be detectable by the dye. As mentioned before, migration time differences between plain DNA and hybrid should decrease with rising DNA length. The reason for this

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

difference is the lower impact of the PNA on the charge-to-hydrodynamic radius ratio with increasing target size. The results presented in Figure 2A showed that heteroduplex formation took place for all oligonucleotide sizes (black lines; hybrid peak) and that the excess DNA was detected in a separate peak when dye was added to the BGE (green lines; DNA peak). As expected, the migration time difference between DNA and hybrid decreased with rising nucleotide count in CE.

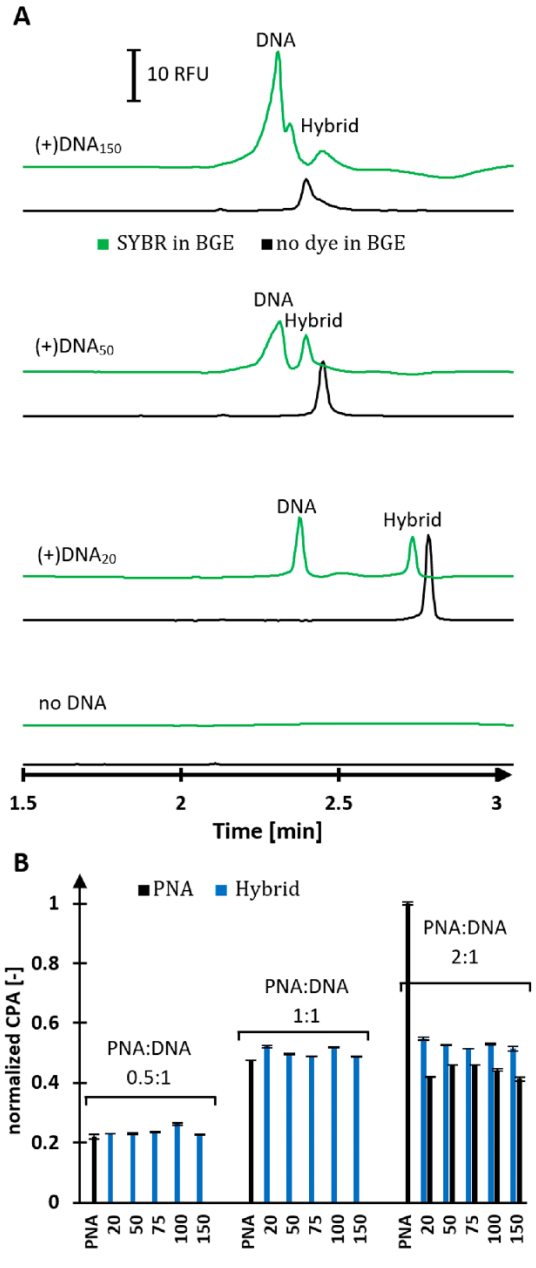


Figure 2. Hybridization of unlabeled ssDNA and labeled PNA oligonucleotides analyzed by CE. (A) $(-)PNA_{15;F}$ and 0−150 nt (+)DNA were injected as a mixture at a concentration of 37.5 nM (PNA) and 75 nM (DNA) each. Samples were measured in a BGE containing SYBR Green II (green) or no dye (black). (B) (+)DNA with varying length was mixed with different amounts of $(-)PNA_{15;F}$ in order to evaluate the potential effects of oligonucleotide size. DNA concentration was stable at 75 nM for all measurements (except PNA-positive control), and PNA varied from 37.5 to 75 nM and 150 nM. Black bars represent the measured CPA of the PNA peak in an electropherogram, while blue bars correspond to the hybrid peak.

In a second step, we wanted to observe the binding efficiency of PNA depending on ssDNA length. For this, the oligonucleotide concentration was kept stable at 75 nM, and the $(-)PNA_{15;F}$ concentration varied with a factor of 0.5× to 2× to DNA. Results presented in Figure 2B showed that in conditions of DNA excess and equimolar concentration, all PNA was bound by the target DNA, no matter what size the oligonucleotide had. In the runs containing a 2-fold PNA excess, half of the PNA was bound to the DNA, the other half was free, and the hybrid peaks were approximately the same size as in the equimolar mixture. This confirms that the PNA is bound quantitatively, which makes this approach not only valuable for identity confirmation but also for quantification.

Interaction of PNA and dsDNA. Using PNA as a probe for dsDNA would be of particular interest for a large variety of therapeutics like dsDNA viruses as adenovirus or herpes virus. As discussed in the previous section, we continued our assessment with fluorescently labeled 20 nt oligonucleotide probes to understand the basic principles and then moved onto longer unlabeled samples. The same $(-)PNA_{15;F}$ probe was used, which binds to the $(+)DNA_{20;F}$ strand and should not interact with the antisense $(-)DNA_{20;F}$ strand. Figure 3A shows the mixture of sense and antisense 20 nt oligonucleotides with different amounts of PNA. The analysis conducted in the 6.6× TBE BGE was able to distinguish between the ssDNA strands and dsDNA (Figure 3A, lower three lines). The reason for this is the more compact structure of the dsDNA double helix compared to more bulky secondary structures of ssDNA. Consequently, the charge-to-hydrodynamic radius ratio was increased for dsDNA in comparison to ssDNA, which led to an accelerated mobility. Another observation was that hybridization of the two strands did not seem to be complete with chosen conditions, even when no PNA was added. This could be due to the low salt concentration in the sample, which was chosen to favor PNA hybridization. When DNA was still in 2-fold excess, the PNA primarily hybridized with the unbound $(+)DNA_{20;F}$. Therefore, the $(\pm)DNA_{20;F}$ peak barely decreased, but a new $(\pm)D/PNA_{20;F}$ heteroduplex appeared, which had a decelerated mobility as was also seen in Figure 1. However, as soon as enough PNA was available (equimolar or excess), $(+)DNA_{20;F}$ was bound completely. The $(\pm)DNA_{20;F}$ peak disappeared and the signal for $(-)DNA_{20;F}$ increased due to $(\pm)D/PNA_{20;F}$ hybrid peak formation (Figure 3A, upper two lines). This showed that the same principles that were demonstrated for ssDNA can be applied to dsDNA. In the next step, we proceeded with the unlabeled and longer targets that were also used in the previous section, but now were complemented with an antisense strand. In Figure 3B, unlabeled 20 and 50 nt DNA at 75 nM were combined with a 1.5-fold PNA excess. For both sizes, the (−) strand showed no formation of hybrid (upper black lines), while the (+) strand was forming hybrids with the $(-)PNA_{15;F}$ probe (green lines). The $(\pm)DNA_{20}$ run was forming hybrids in the same amount as the $(+)DNA_{20}$ alone, which reproduced the results with labeled samples from Figure 3A. However, for a longer $(\pm)DNA_{50}$, the picture changed. The hybrid peak was significantly reduced, and the PNA peak indicated that $(+)DNA_{50}$ was not bound completely to the PNA but was hybridizing $(-)DNA_{50}$ (Figure 3B, upper two lines) A significantly higher PNA excess of 20× did not change this pattern (data not shown). To get a better picture of this

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

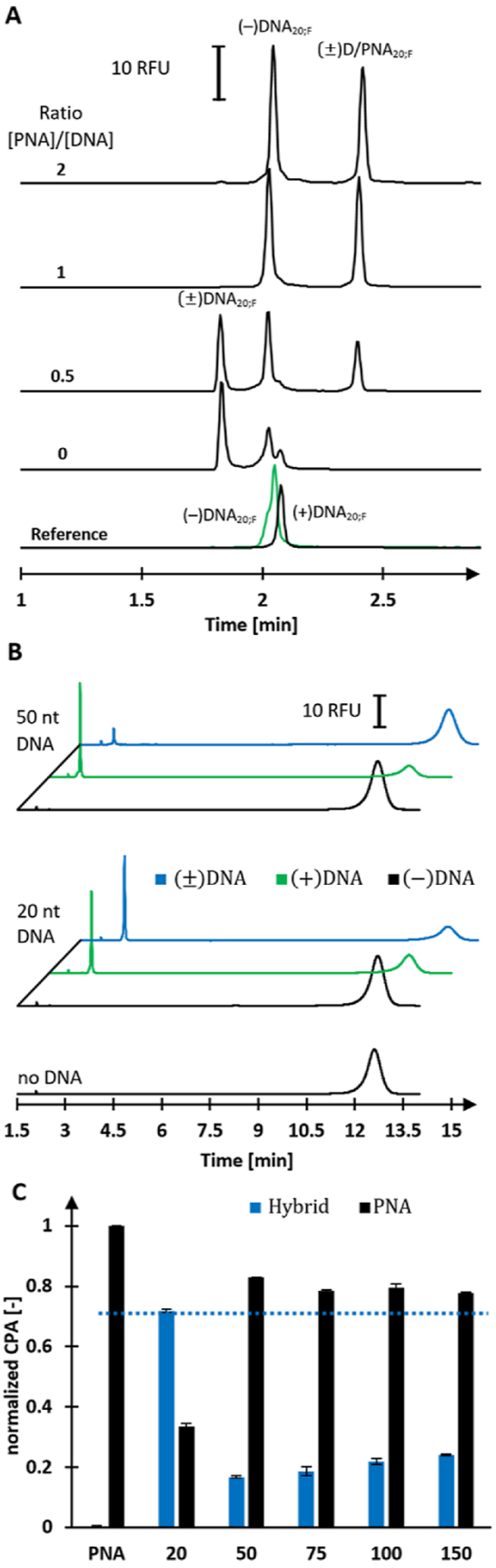


Figure 3. Competitive hybridization of DNA and PNA oligonucleotides analyzed by CE. (A) Two labeled and complementary 20 nt oligonucleotides were mixed 1:1 and different concentrations of PNA were added. ssDNA was always at 150 nM, while the PNA varied with a multiplier from 0 to 2. The "Reference" line is an overlay of the two ssDNA strands injected independently. In the "0" line, both complementary strands are actual mixes, but no PNA is added. For lines "0.5", "1", and "2", the corresponding relative amount of PNA was added. (B) Two unlabeled and complementary 20 nt and 50 nt oligonucleotides were mixed 1:1 at 75 nM, and a 1.5-fold excess of PNA was added (blue lines). The results were compared with a mixture of PNA and a ssDNA oligonucleotide. Either the (+) strand that is binding to the PNA (green lines) or the (−) strand that serves as a negative control (black lines). The (+)DNA lines are shifted by 1 min, while the dsDNA lines are shifted by 2 min, respectively. (C) Comparison of CPAs of the PNA peak to the hybrid peak in competitive conditions described in (B) for dsDNA oligos with a length varying from 20 to 150 nt. The dashed line indicates the expected CPA if all of the target DNA would be bound by the PNA (as is the case for the 20 nt strands).

effect, larger DNAs were tested. It was found that the longer the target dsDNA was, the broader the peak became. This showed that the hybrid was not completely stable. When analyzing the CPAs of these hybrid peaks, it was observed that only 25−35% of the target DNA could be bound by the PNA (Figure 3C, blue bars vs dashed line). These results showed that PNA was not able to completely displace a longer DNA strand that binds to the same sequence. For this reason, PNA hybridizations can be useful for quantitative analysis of very short dsDNA fragments that are similar in size to the used PNA.

Application of the Technique in Complex Test Sample Matrices. In the previous section, we demonstrated that ssDNA binding to PNA probes is very strong and, thus, quantification of target sequences was possible. In order to find a working range, we prepared mixtures of 1000 nM PNA with a 100 nt DNA. The DNA concentration varied from 12.5 to 800 nM, and the change of hybrid CPA was measured (data not shown). Since the results showed a strong linear response (R^2 = 0.9995), we decided to test if the method can be used to quantify a specific target in the presence of an unspecific DNA at high concentration levels. Therefore, a mixture of unspecific 180 nt DNA ((X)DNA_{180}) at 1000 nM with a target (+)DNA_{100} with varying concentration was prepared. The results are presented in Figure 4A. The mixture of the (+)DNA_{100} and (X)DNA_{180} strands (320 and 1000 nM, respectively) were injected in a BGE-containing SYBR and compared with a simple (+)DNA_{100} (320 nM) injection. Although the exact interaction of SYBR green II and ssNAs is still not fully understood, it is hypothesized that it binds to nucleotides in a regular manner, similarly to intercalating dyes. If this would be the case, the signal intensity of the DNA signal (Figure 4A, upper line) should be approx. 15% of the signal related to the mixed DNA (lower line). Note that not only the DNA concentration is different but also the 180 nt DNA induces a higher signal intensity due to staining. However, the observed signal was <9% most likely because SYBR Green II tends to bind with a different affinity to the individual nucleotides as more recent research has shown.[35] When 500 nM PNA was added to these samples, and the measurement was performed without a dye, the signals remained unchanged (Figure 4B). This showed that PNAs are very sequence specific, even when a large amount of unspecific DNA is present in the sample. To test the linearity of the method

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

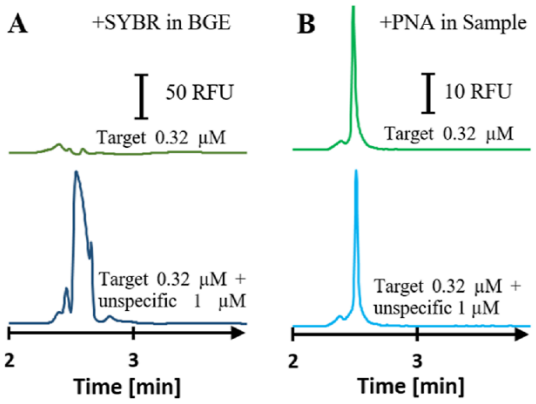


Figure 4. Quantification of a 100 nt DNA sequence in a mixed DNA sample. (A) Limitation of SYBR Green for quantification: (+)DNA_{100} (target) was measured individually at 0.32 μM (upper line) or mixed with 1 μM of unspecific (X)DNA_{180} (lower line) in a BGE containing SYBR Green II. (B) To test the specificity of the method, samples from (A) were mixed with 500 nM (−)$PNA_{15;F}$ (probe) and measured in a BGE that does not contain SYBR Green.

under even more unfavorable conditions, the target (+)-DNA_{100} was prepared in various concentrations from 10 to 320 nM. Additionally, an unspecific (X)DNA_{180} was added at a concentration of 1000 nM. As a result, the target varied from 1 to 24.2% of the total DNA strand molarity. The obtained linearity is satisfactory (R^2 = 0.9991; Figure S3) and confirms that the test method is specific, selective, and robust to quantify traces of impurities in oligonucleotide samples.

Plasmids and Corresponding Gene Therapy Products. Our initial goal was to identify DNA species in gene therapy products. However, we showed that binding of dsDNA by PNA probes was only partial, unless the target had a similar size as the PNA. PNA synthesis is limited to approx. 35 nt, and the target sequences exceeded this size in almost all cases. Hence, with a single PNA probe, only qualitative analysis of dsDNA samples was possible. To overcome this issue, we decided to use multiple PNA probes that should hinder the target DNA-forming dsDNA. An additional advantage of this approach is a further increase of the signal intensity since multiple labels are binding the target. The principle was tested using (±)DNA_{75} oligonucleotides shown in Figure 3C. (±)DNA_{75} oligonucleotides were hybridized with the (−)-$PNA_{15;F}$ strand and two additional (−)$PNA_{20;F}$ that bind upstream and downstream of the initial target sequence. Thereby, only 20 nt of the target DNA was not covered. This should prevent the formation of dsDNA and favor the formation of the DNA$(PNA)_3$ quadruplex. Indeed, under these conditions, the target DNA was bound completely and remains unchanged regarding signal intensity, when compared with a (+)DNA_{75} run (Figure S4).

With this demonstration, we moved to a more complex system: the rAAV production plasmid (containing the transgene) and the corresponding rAAV. Although rAAVs have a ssDNA genome, there is a mixture of (+) and (−) polarity DNA encapsidated, so dsDNA starts to form as soon as the DNA gets extracted.[36,37] In order to get fragments that are small enough, a digestion was needed. For a proof of concept, we designed a digest that cuts a sequence of 54 nt as a target, which is present in the EGFP gene. The (−)$PNA_{15;F}$ and one of the (−)$PNA_{20;F}$ were used as probes. Because of overhangs that were produced by the restriction enzymes, only 46 nt of the target was in fact double-stranded, of which 35 nt was covered by PNAs. Before digestion, the extracted rAAV DNA and plasmid DNA were measured with a UV photometer in TE buffer. After digestion, all samples were hybridized with both PNAs each at 75 nM. Figure 5 shows the resulting hybrid

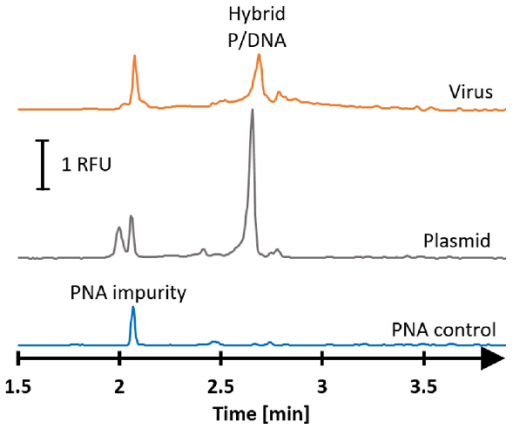


Figure 5. Hybridization of digested dsDNA using two PNAs. (−)$PNA_{15;F}$ and (−)$PNA_{20;F}$ were each 75 nM and mixed with digested rAAV production plasmid (middle line) and digested rAAV genomic DNA (upper line). The target sequence was in the EGFP gene and had a length of 54 nt.

peaks in comparison to a PNA control (lower line). PNA impurities led to a peak close to 2 min, while the target DNA was detected for the plasmid (middle line) and virus (upper line) around 2.6 min. The relative CPA percentages of these hybrid peaks in comparison to the total CPA, which includes the PNA signals, were linear (R^2 > 0.99) for the plasmid and viral DNA. Assuming that the injected amount of 150 nM PNA results in a corresponding signal, the concentration of DNA in the test sample solution (final concentration) can be deduced from the CPA percentage of the hybrid peak. Considering all dilution and purification steps, test sample concentrations of 4.9 nM (plasmid dsDNA) and 3.0 nM (virus ssDNA) were calculated (exact calculations can be found in the Supporting Information). For the plasmid, this value was very close to the photometric measurement of 4.6 nM, while the calculated virus DNA concentration was approximately half the UV measurement of 6.4 nM. To double-check this result, we compared the value we obtained for the virus with quantitative PCR (qPCR) results provided by the manufacturer (1.03×10^{13} vg/mL). For exact comparability, a factor needs to be applied on the initial value obtained by CE: the method detects only (+)DNA, while qPCR measures all genomes, sense, and antisense. A comprehensive comparison of the values obtained by various techniques and their conversion in different units is summarized in Table 1. Under the assumption that every genome is intact, a concentration of 17.1 nM was measured by qPCR, so our CE method detected only 17.5% of this value, while UV found 36.9%. However, qPCR was performed using primers for the inverted terminal repeat (ITRs) sequence, which serves as the primary packaging signal of the DNA within the vector plasmid during the manufacturing process. It is thus possible that fragmented sequences that did not contain the EGFP target were packaged in the capsids. Additionally, the incorporated plasmid backbone also carries ITRs which would lead to an increased value for qPCR. A report from Burnham et al. found

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

Table 1. Comparison of DNA Concentration Values That Were Detected with Different Techniques for a rAAV Sample[a]

	titer [vg/mL]	ρ [ng/μL]	c [nM]	standard deviation (%)
qPCR	1.03 × 1013[b]	13.2	17.1	±10[c]
UV	3.9 × 1012	4.9[b]	6.4	±2.2
CE	1.8 × 1012	2.3	3.0[b]	±1.5

[a]The exact calculations can be found in the Supporting Information. [b]Indicates that this value was obtained by actual measurement. All other values in the corresponding row are calculated out of this value under the assumption that each technique detects only intact ssDNA genomes with a size of 2500 nt and M_w of 770 kg/mol. [c]This value was not obtained through measurement but from the literature.[19]

that under unfavorable conditions, more than 80% of rAAVs may not contain the intended genome.[38] For verification, this rAAV sample was also measured with CGE, and we found plenty of unknown and fragmented species (see Figure S5). A transduction efficiency test conducted on U87 cells showed only limited production of EGFP when compared to other rAAV batches (data not shown). The picture that is drawn by combining qPCR, UV photometry, PNA-CE assay, CGE, and potency assays could be an indication that only a very small fraction of the analyzed rAAV sample contained an EGFP target sequence that was bound by PNA probes and highlights that qPCR assays that use ITRs for quantification might overestimate the actual titer. Our CE approach proved to be valuable for this highly complex sample. However, the use of purification columns brings the risk of sample loss, and the choice of DNA nucleases must be individual for each sequence that needs to be detected. The nucleases digest only dsDNA, so rAAV reannealing must be complete for an accurate estimation by CE.

Although CE and qPCR provide some sequence information, the data can be misleading when a target region is chosen, which is overrepresented in the capsids, for example, the ITRs. It remains difficult to assess if a target is a good choice due to potential fragmentation.

Quantification of (m)RNA and Determination of the Limit of Quantification. Most RNA-containing medications are single-stranded, which is an advantage for the CE method. Digests are considered unnecessary since no competitive strand is present, and intramolecular interactions should be a less critical factor than dsNA formation. We applied the PNA hybridization method for a mRNA that was coding for EGFP and had a size of approx. 996 nt ((+)RNA_{996}), including the poly A tail, according to the manufacturer. To this end, (−)$PNA_{15;F}$ used in previous experiments and two additional (−)$PNA_{20;F}$ to enhance the signal intensity were chosen as probes. The obtained hybrid signal was in a range of 0−50 nM (R^2 = 0.994). For limit of quantification (LOQ) estimation, we used the same (+)RNA_{996} and three corresponding PNAs. Each of them was set at 20 nM, and the initial RNA concentration was set to 15 nM (Figure 6, upper line) and then diluted to get a final concentration of 0.94 nM (middle line). All conditions showed hybrid formation with the corresponding signal intensity. In our experiment, the limitations for the LOQ resulted from the nature of the PNA probes, and we did not reach the device maximum. PNAs do contain some impurities themselves that are typically below 1% of CPA of the main peak. When using a large excess of PNA, these signals might comigrate with the hybrid signal, as

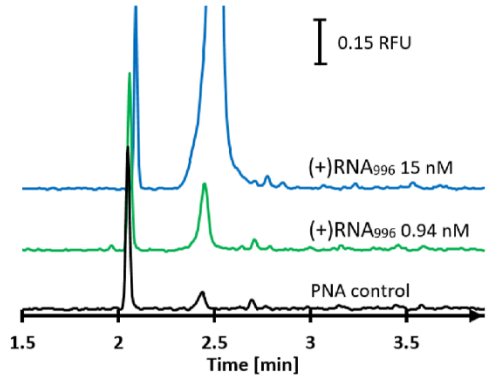


Figure 6. Hybridization of three PNAs and mRNA and estimation of the LOQ. (−)$PNA_{15;F}$ and two (−)$PNA_{20;F}$ were each 20 nM and mixed with mRNA (+)RNA_{996} at 15 nM (upper line) and 0.9375 nM (middle line).

can be seen in Figure 6. However, this problem can be circumvented by using a lower PNA excess and thus decreasing the PNA impurity peaks. Considering all these points, the LOQ under carefully chosen conditions can certainly be in the picomolar range. Direct detection of mRNA is an especially promising application of the CE method, as it avoids the reverse transcriptase step that qPCR requires and therefore saves time and reduces biases. PCR methods cannot be outpaced in terms of sensitivity since they are theoretically able to detect a single molecule. Nevertheless, the demonstrated sensitivity of the CE method is comfortably in the range of most RNA therapeutics such as mRNA vaccines.

■ CONCLUSIONS

The space of advanced therapeutic modalities is currently evolving in rapid pace necessitating continuous evolvement of analytical quality control methods. Quantification and identification of nucleic acid species in these drug products are of pivotal importance to ensure product quality. We presented a versatile CE-based gel-free hybridization assay that uses fluorescently labeled PNAs as probes for NA identity. It is applicable for a wide size range of NAs and, therefore, useful for drug product formulations with different levels of complexity such as oligonucleotides, mRNA vaccines, or rAAVs. Suitable application fields and considerations that need to be made before starting method development are summarized in Figure 7. For ssNAs (Figure 7A,B) ranging from 20 to 1000 nt, the method is an excellent choice that provides information on quantity and identity in a broad range with high reproducibility. It proved to be highly specific by detecting ssDNA oligonucleotide traces in a complex test sample matrix as was also described previously.[24] In the present work, we were able to apply this concept for mRNA, while having an LOQ in the picomolar range when multiple probes were used. For double-stranded samples (Figure 7C,D), only DNA fragments that are similar to PNA size can be quantified. This limitation can be circumvented by digesting and by using multiple probes. By means of digesting a transfer plasmid and the corresponding rAAV genomic DNA, we have shown in a concept study that quantification of complex samples without a standard is possible, but molecule-specific optimization and calibration with an appropriate reference material might be advantageous. The comparison with orthogonal techniques showed that none of the methods are

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

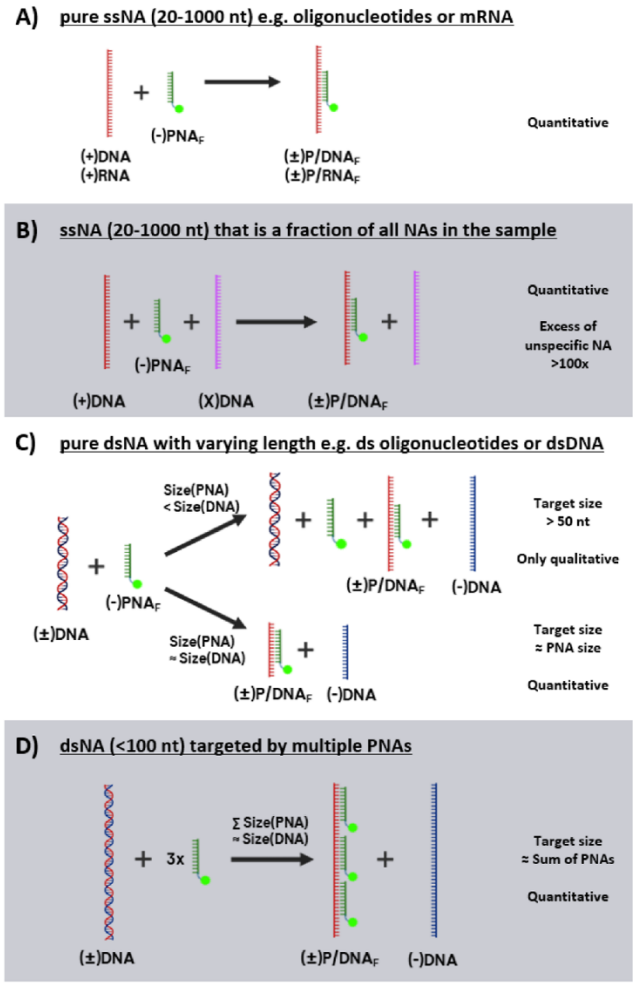


Figure 7. Application fields of a CE-based PNA assay.

fully capable of identifying and thus quantifying NAs for this particularly complex sample. For the purpose of quality control, several techniques must be considered in combination to gain a clear picture of complex gene therapy analytes. In this regard, the use of PNAs as hybridization probes in CE is a powerful orthogonal approach to qPCR adding significant value to the analytical toolbox of nucleic acid therapeutics.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c04813.

Addendum to the experimental section; simplified depiction of PNA and DNA hybridization; details on chemical agents used for optimization of the assay; expansion of Figure S1; linearity estimation for Figure S4; proof of concept of DNA/PNA multiplex formation; CGE analysis of the rAAV sample; and exact calculations for Table 1 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Andrei Hutanu − *Analytical Development and Quality Control, Pharma Technical Development Europe, F. Hoffmann-La Roche AG, Basel 4070, Switzerland; University of Basel, Basel 4056, Switzerland;* orcid.org/0000-0003-0174-5250; Email: a.hutanu@unibas.ch

Authors

Chiara Signori − *Analytical Development and Quality Control, Pharma Technical Development Europe, F. Hoffmann-La Roche AG, Basel 4070, Switzerland*

Bernd Moritz − *Analytical Development and Quality Control, Pharma Technical Development Europe, F. Hoffmann-La Roche AG, Basel 4070, Switzerland*

Manuel Gregoritza − *Analytical Development and Quality Control, Pharma Technical Development Europe, F. Hoffmann-La Roche AG, Basel 4070, Switzerland*

Adelheid Rohde − *Analytical Development and Quality Control, Pharma Technical Development Europe, F. Hoffmann-La Roche AG, Basel 4070, Switzerland*

Maria A. Schwarz − *University of Basel, Basel 4056, Switzerland; Solvias AG, Kaiseraugst 4303, Switzerland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c04813

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors wish to thank Dimitri Maier for conducting a potency assay of the rAAV sample discussed in this paper. The authors acknowledge the web tool BioRender.com, which was used to create the abstract figure, Figure 7, and S1.

■ REFERENCES

(1) Ferreira, C. R. *Am. J. Med. Genet.* **2019**, *179*, 885−892.

(2) Haendel, M.; Vasilevsky, N.; Unni, D.; Bologa, C.; Harris, N.; Rehm, H.; Hamosh, A.; Baynam, G.; Groza, T.; McMurry, J.; Dawkins, H.; Rath, A.; Thaxon, C.; Bocci, G.; Joachimiak, M. P.; Köhler, S.; Robinson, P. N.; Mungall, C.; Oprea, T. I. *Nat. Rev. Drug Discovery* **2020**, *19*, 77−78.

(3) Keeler, A. M.; Flotte, T. R. *Annu. Rev. Virol.* **2019**, *6*, 601−621.

(4) Wang, D.; Tai, P. W.; Gao, G. *Nat. Rev. Drug Discov.* **2019**, *18*, 358−378.

(5) Bulaklak, K.; Gersbach, C. A. *Nat. Commun.* **2020**, *11*, 1−4.

(6) Brooks, P. J.; Ottinger, E. A.; Portero, D.; Lomash, R. M.; Alimardanov, A.; Terse, P.; Xu, X.; Chandler, R. J.; Hauserman, J. G.; Esposito, E.; Bönnemann, C. G.; Venditti, C. P.; Austin, C. P.; Pariser, A.; Lo, D. C.; Averion, G. V.; Balakrishnan, K.; Burden, S. J.; Campbell, E.; Chen, C.; Driscoll, C.; Dukhanina, O.; Ferry, S.; Foley, A. R.; Li, L.; Manoli, I.; Mendoza, C.; Oury, J.; Porter, F. D.; Portilla, L.; Saade, D.; Shchelochkov, O. A.; Sloan, J. L.; Todd, J.; Toney, L.; Ryzin, C. V.; Vepa, S.; Wagner, E.; Wang, A. *Hum. Gene Ther.* **2020**, *31*, 1034−1042.

(7) Colella, P.; Ronzitti, G.; Mingozzi, F. *Mol. Ther.–Methods Clin. Dev.* **2018**, *8*, 87−104.

(8) Ghosh, S.; Brown, A. M.; Jenkins, C.; Campbell, K. *Appl. Biosaf.* **2020**, *25*, 7−18.

(9) Nidetz, N. F.; McGee, M. C.; Tse, V. T.; Li, C.; Cong, L.; Li, Y.; Huang, W. *Pharmacol. Ther.* **2020**, *207*, 107453.

(10) Papanikolaou, E.; Bosio, A. *Frontiers in Genome Ed.* **2021**, *3*, 618346.

(11) Gimpel, A. L.; Katsikis, G.; Sha, S.; Maloney, A. J.; Hong, M. S.; Nguyen, T. N.; Wolfrum, J.; Springs, S. L.; Sinskey, A. J.; Manalis, S. R. *Mol. Ther.–Methods Clin. Dev.* **2021**, *20*, 740−754.

(12) Kassim, S. H. *Cell Gene Ther. Insights* **2017**, *3*, 227−223.
(13) Schmidt, A.; Helgers, H.; Vetter, F. L.; Zobel-Roos, S.; Hengelbrock, A.; Strube, J. *Processes* **2022**, *10*, 1783.
(14) Wright, J. F. *Biotechnol. J.* **2021**, *16*, 2000022.
(15) Armstrong, M.; Jonscher, K.; Reisdorph, N. A. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2717−2726.
(16) Fekete, S.; Doneanu, C.; Addepalli, B.; Gaye, M.; Nguyen, J.; Alden, B.; Birdsall, R.; Han, D.; Isaac, G.; Lauber, M. *J. Pharm. Biomed. Anal.* **2022**, *224*, 115174.
(17) Wei, B.; Goyon, A.; Zhang, K. *J. Pharm. Biomed. Anal.* **2022**, *219*, 114928.
(18) Barczak, W.; Suchorska, W.; Rubiś, B.; Kulcenty, K. *Mol. Biotechnol.* **2015**, *57*, 195−200.
(19) Aurnhammer, C.; Haase, M.; Muether, N.; Hausl, M.; Rauschhuber, C.; Huber, I.; Nitschko, H.; Busch, U.; Sing, A.; Ehrhardt, A. *Hum. Gene Ther. Methods* **2012**, *23*, 18−28.
(20) Baldwin, J.; Piplani, S.; Sakala, I. G.; Honda-Okubo, Y.; Li, L.; Petrovsky, N. *Bioanalysis* **2021**, *13*, 1805−1826.
(21) Guerin, K.; Rego, M.; Bourges, D.; Ersing, I.; Haery, L.; Harten DeMaio, K.; Sanders, E.; Tasissa, M.; Kostman, M.; Tillgren, M. *Hum. Gene Ther.* **2020**, *31*, 664−678.
(22) Sanyal, G.; Särnefält, A.; Kumar, A. *npj Vaccines* **2021**, *6*, 1−9.
(23) Cring, M. R.; Sheffield, V. C. *Gene Ther.* **2022**, *29*, 3−12.
(24) Hu, L.; Anand, M.; Krylova, S. M.; Yang, B. B.; Liu, S. K.; Yousef, G. M.; Krylov, S. N. *Anal. Chem.* **2018**, *90*, 14610−14615.
(25) Gupta, A.; Mishra, A.; Puri, N. *J. Biotechnol.* **2017**, *259*, 148−159.
(26) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M. *Nature* **1994**, *368*, 561−563.
(27) Basile, A.; Giuliani, A.; Pirri, G.; Chiari, M. *Electrophoresis* **2002**, *23*, 926−929.
(28) Hu, L.; Krylova, S. M.; Liu, S. K.; Yousef, G. M.; Krylov, S. N. *Anal. Chem.* **2020**, *92*, 14251−14258.
(29) Perry-O'Keefe, H.; Yao, X. W.; Coull, J. M.; Fuchs, M.; Egholm, M. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 14670−14675.
(30) Rose, D. J. *Anal. Chem.* **1993**, *65*, 3545−3549.
(31) Tedeschi, T.; Chiari, M.; Galaverna, G.; Sforza, S.; Cretich, M.; Corradini, R.; Marchelli, R. *Electrophoresis* **2005**, *26*, 4310−4316.
(32) Tsukada, H.; Kundu, L. M.; Matsuoka, Y.; Kanayama, N.; Takarada, T.; Maeda, M. *Anal. Biochem.* **2013**, *433*, 150−152.
(33) Hutanu, A.; Boelsterli, D.; Schmidli, C.; Montealegre, C.; Thai, M. H. D.; Bobaly, B.; Koch, M.; Schwarz, M. A. *Electrophoresis* **2022**, *43*, 1107−1117.
(34) Rose, K.; Mason, J. O.; Lathe, R. *BioTechniques* **2002**, *33*, 54−58.
(35) Zhang, B.; Jiang, J.; Yuan, Y.; Guan, Y. *Chem. Res. Chin. Univ.* **2019**, *35*, 1119−1123.
(36) Gonçalves, M. A. *Virol. J.* **2005**, *2*, 1−17.
(37) Mayor, H. D.; Torikai, K.; Melnick, J. L.; Mandel, M. *Science* **1969**, *166*, 1280−1282.
(38) Burnham, B.; Nass, S.; Kong, E.; Mattingly, M.; Woodcock, D.; Song, A.; Wadsworth, S.; Cheng, S. H.; Scaria, A.; O'Riordan, C. R. *Hum. Gene Ther. Methods* **2015**, *26*, 228.

https://doi.org/10.1021/acs.analchem.2c04813
Anal. Chem. 2023, 95, 4914−4922

Supporting Information

Using Peptide Nucleic Acid Hybridization Probes for Qualitative and Quantitative Analysis of Nucleic Acid Therapeutics by Capillary Electrophoresis

Analytical Chemistry. 2023 Mar; 95(11):4914-22..

1 Sequence of nucleic acids used in this study

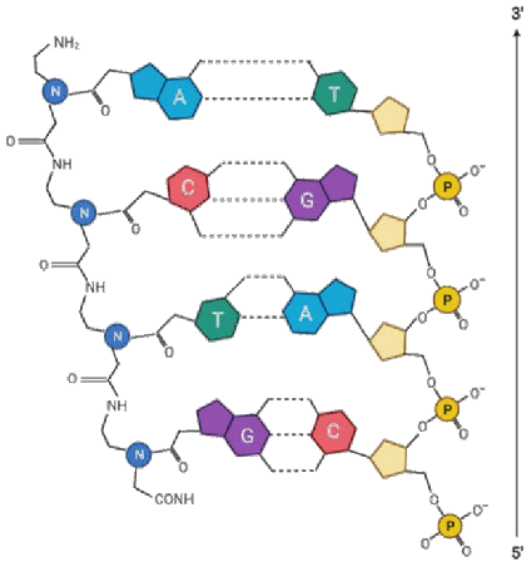
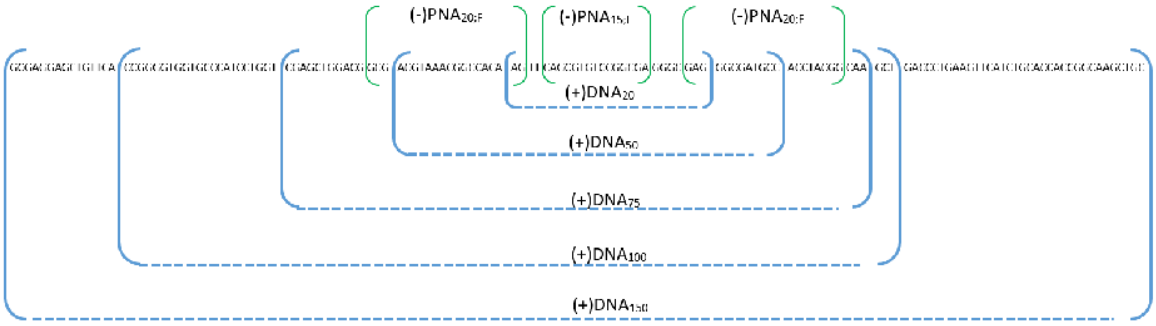


Figure S1: Simplified depiction of PNA and DNA forming a hybrid.

All following sequences are written in a 5′-to-3′ direction.



Scheme S1: Used DNA oligonucleotides and their target sequence bound by PNAs

Table S1: Sequence of utilized DNA oligomers in this study. Only DNA sense (+) strands are shown; antisense (-) were the reverse complement; (+)DNA_{20} was also labeled with FAM (+)$DNA_{20;F}$ depending on the experiment; (x)DNA_{180} was a random sequence and no antiparallel strand was used.

Name	Sequence
(+)DNA_{20}	AGTTCAGCGTGTCCGGCGAG
(+)DNA_{50}	ACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
(+)DNA_{75}	CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGA TGCCACCTACGGCAA
(+)DNA_{100}	CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGT GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT
(+)DNA_{150}	GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTG AAGTTCATCTGCACCACCGGCAAGCTGC
(x) DNA_{180}	GGGTCCCCGTACGCCCACGCCTGCTCGCCCGTCGCATGCATAAAAGGGTAGACGCGATGACG GGCGCACTCCTCTGGCCACATCGGTTCCTGCTCCCGCTGGGATCCAGGGTTGGCGGCCGAGG CCGCCTCTCCGTAGTGAGTCCTTCGTCCGTGGCTAACCGTGCCAGATCGCCTGG

Table S2: Sequence of utilized PNA oligomers in this study.

Name	Comment	Sequence
(-)$PNA_{15;F}$	Used in all experiments	FAM–OO–TCGCCGGACACGCTG
(X)$PNA_{15;F}$	Unspecific	FAM–OO–ATTTATACCCTGCGC
(-)$PNA_{20;F}$	downstream of (-)$PNA_{15;F}$	FAM–OO–CTTGTGGCCGTTTACGTCGC
(-)$PNA_{20;F}$	upstream of (-)$PNA_{15;F}$	FAM–OO–CCGTAGGTGGCATCGCCCTC
(-)$PNA_{20;F}$	Used for mRNA	FAM–OO–TCTCGTTGGGGTCTTTGCTC-Glu-Glu
(-)$PNA_{20;F}$	Used for mRNA	FAM–OO–TTCTGCTTGTCGGCCATGAT-Glu-Glu

The complete sequence of the EGFP gene (green markup was targeted by PNAs)

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTT
CAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG
TGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCCTGGGGCACAAGCTGGAGTACAACTACAACA
GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCA
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA
CCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTC
TCGGCATGGACGAGCTGTACAAGTAA

After digest with BtgZI and EaeI the following sequence of EGFP was the target for two PNAs:

GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG

2 Addendum to the experimental section

Reagents

NaOH (Cat. no. 28244.295) was from VWR Chemicals (Radnor; USA). HCl (Cat. no. 387800010) was purchased from Acros organics (Geel; Belgium) and 10% SDS Solution (Cat. no. 1610416) from Bio-Rad Laboratories (Hercules; USA). The following chemicals were from Sigma-Aldrich/Merck KGaA (Darmstadt; Germany): Acetonitrile (Cat. no. 34851) DMSO (Cat. no. 276855), 0.5 M EDTA (Cat. no. 15575020), formamide (Cat. no. 47671), cetyltrimethylammonium bromide (CTAB; Cat. no. H6269), polyvinylpyrrolidone (PVP; Cat. no. 437190), 10x Tris borate EDTA (TBE) buffer (Cat. no. 574795), Tris Base (Cat. no. T6066), 10% Tween® 20 solution (Polysorbate 20; Cat. no. 11332465001) and urea (Cat. no. U6504). SYBR™ Green II RNA Gel Stain 10,000 x concentrate in DMSO (Cat. no. S7564), UltraPure™ DNase/RNase-Free Distilled Water (Cat. no. 10977015) were purchased from Thermo Fisher Scientific (Waltham; USA). The QIAquick PCR purification kit (Cat. no. 28104) and QIAquick nucleotide removal kit (Cat. no. 28306) used for DNA purification were from Qiagen (Germantown, USA). EaeI (Cat. no. R0508) and BtgZI (Cat. no. R0703) nucleases were obtained from New England Biolabs Inc. (Ipswich, USA). Water of HPLC grade was prepared in a Milli-Q-Station (Merck Millipore/Merck KGaA; Darmstadt; Germany). Solutions were filtered through 0.2 µm membrane filters (Corning; New York; USA).

Digestion

Prior to digestion 125 µL rAAV genome was purified with the Qiaquick PCR purification kit as described in the manufacturers protocol or our previous report[24] and eluted in 100 µL TE buffer. Afterwards the sample was heated at 95 °C for 10 min and cooled down slowly for 90 min in the heat block to allow hybridization. DNA concentration was measured with a NanoDrop™ 2000 spectral photometer (Cat. no. ND-2000) from Thermo Fisher Scientific (Waltham, USA). The plasmid stock was diluted in TE buffer to the same molar concentration and then the concentration was checked with the photometer. The EaeI digest was performed in a total volume of 119 µL containing 30 U (6 µL) of EaeI, 12.5 µL digestion buffer (rCutSmart 10x; provided with the nuclease) and 0.6 µg (virus) or 1.4 µg (plasmid) of DNA at 37 °C for 60 min. Afterwards 30 U (6 µL) of BtgZI were added to a total volume of 125 µL and the samples were incubated at 60 °C for 60 min. Purification of the digested DNA was performed with the Qiaquick nucleotide removal kit as described in the manufacturers manual and samples were eluted in 30 µL TE buffer. A twofold dilution row of the samples was created (four concentrations in total) to check for linearity of the hybrid signal. The different dilutions were then hybridized with target PNAs at 75 nM as described above and analyzed by CE.

CGE

The CGE control was performed as described in our previous report[33] in a gel buffer consisting out of 2% PVP, 4 M urea in 1xTBE solution with 1:25,000 diluted SYBR Green II at -6 kV in a 20/30 cm, 50 µm I.D. fused silica capillary

3 Chemical agents used for optimization of the assay

Table S3: Agents used for optimization of hybridization reaction

Agent	Concentration Range	Effect
DMSO	0.1-50% (v/v)	PNA: reproducible above 10% DNA: none Hybrid: slight stabilization above 20% Migration time shifts at high concentration for all species
SDS	0.3-290 mM CMC≈7.5 mM	PNA: reproducible at any concentration above 1 mM, peak broadening above CMC DNA: no effect up to 3 mM, peak broadening above CMC Hybrid: sharper peaks above 1 mM, broadening above CMC Best conditions were found at 2 mM
ACN	0.1-30% (v/v)	PNA: reproducible above 10% DNA: none Hybrid: peak broadening Considerable evaporation in autosampler at high concentration
CTAB	0.27 - 137 mM CMC≈0.9 mM	DNA peak broadening at 0.27 mM; visible precipitation at 2.7 mM
Polysorbate 20	0.01-8.4% (w/v) CMC≈0.06%	PNA: reproducible above 0.01%, no impact with higher concentrations DNA: none Hybrid: none; destabilizing at 8.4%
Urea	0.1 - 7 M	PNA: reproducible above 0.1 M, no impact with higher concentrations DNA: peak broadening starting with 1 M Hybrid: destabilizing above 3M

Table S4: BGE optimization screening. 1x TBE was the base and the variables/agents below were altered (e.g. 1X TBE; 30% DMSO).

Agent	Concentration Range	Effect
DMSO	0.1-30% (v/v)	PNA: reproducible above 1% DNA: none Hybrid: slight stabilization above 20% Migration time shifts at high concentration for all species
SDS	0.3-290 mM CMC≈7.5 mM	SDS tends to interfere with the coating of the capillary, results are not reproducible
Urea	0.1 – 7 M (v/v)	PNA: not reproducible DNA: none Hybrid: peak broadening above 5 M
Higher pH	pH 10	Only feasible in fs capillary, PNA not reproducible, Fluorescence signal decreases.
Salt concentration	0.25 – 6.6x TBE	PNA: very reproducible at 6.6x TBE DNA: higher signal intensity at 6.6x TBE Hybrid: Improved separation from DNA peak, correlating with increasing salt concentration

4 Supplementary figures

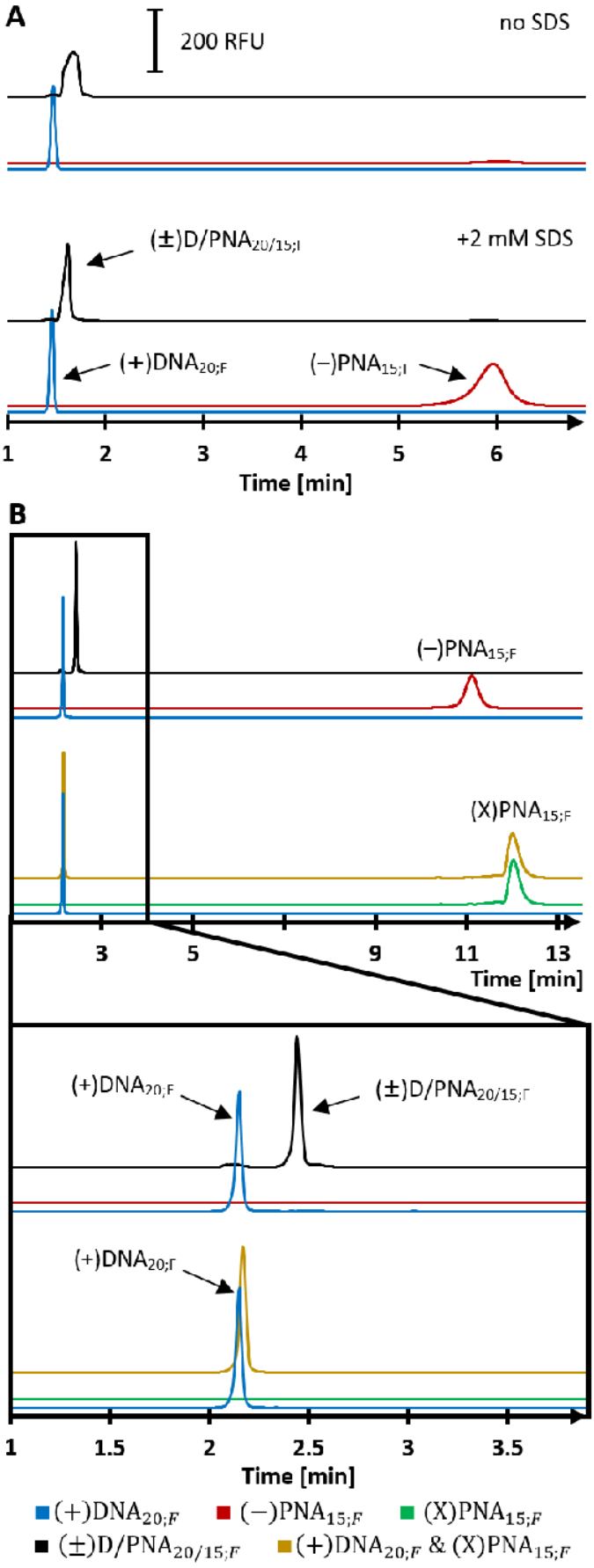


Figure S2: Hybridization of DNA and PNA oligonucleotides analyzed by CZE. The RFU scale is valid for A) and B). PNA and DNA signals are overlaid while mixed samples are shifted on the y-axis for clarity. **A)** Starting point for method development. PNA and DNA samples were injected individually or as a mixture at a concentration of 1 µM each. PNA signals are not reproducible under these conditions while a clear DNA peak is visible (upper red and blue line). However, hybridization can take place and a shift of the DNA signal can be detected when incubated with PNA (upper black line). Addition of 2 mM SDS to the samples before incubation leads to reproducible signals for PNAs, improves peak shape of the hybrid, while it has no negative effect on DNA (lower three lines). BGE: 1x TBE, 3 M urea. **B)** Separation further improves by adjusting the BGE to 6.6x TBE with 3 M urea. Signal intensity and resolution between hybrid and DNA peaks were improved in comparison to A) (upper three lines). The lower three lines serve as negative control: Incubation of DNA with a PNA that has an unspecific sequence ((X)$PNA_{15;F}$) does not lead to a typical migration shift and areas of PNA and DNA peaks are not affected (lower three lines). The insert reprints the same data set from 1-4 min.

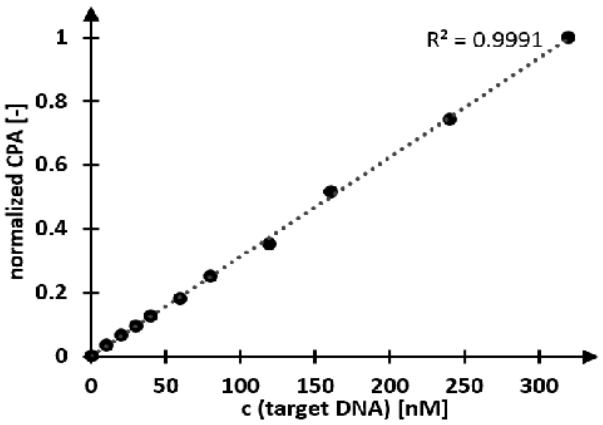


Figure S3: For linearity estimation, (+)DNA_{100} from Figure 4 was varied in concentration (10-320 nM) while it was mixed with 1000 nM of (X)DNA_{180}. (–)$PNA_{15;F}$ was kept constant at 500 nM for all measurements. The target DNA concentration was plotted against the CPA of the hybrid peak that was observed in the electropherogram.

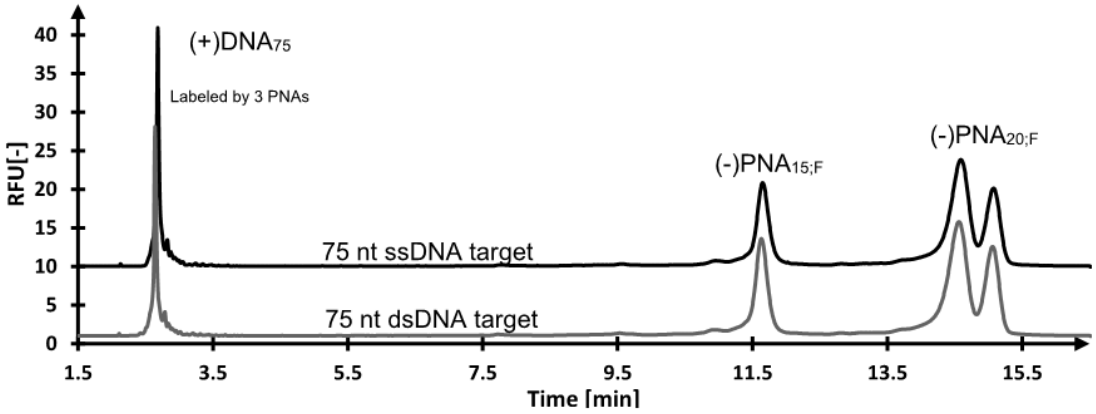


Figure S4: Two unlabeled and complementary 75 nt DNAs were mixed at 75 nM and a 2 fold excess of three PNAs was added: (-)$PNA_{15;F}$ strand and two additional (-)$PNA_{20;F}$ that bind upstream and downstream of the initial target sequence (upper line; dsDNA). The results were compared with a mixture of PNA and the ssDNA oligo that does not contain the (-)DNA_{75} strand (lower line; ssDNA).

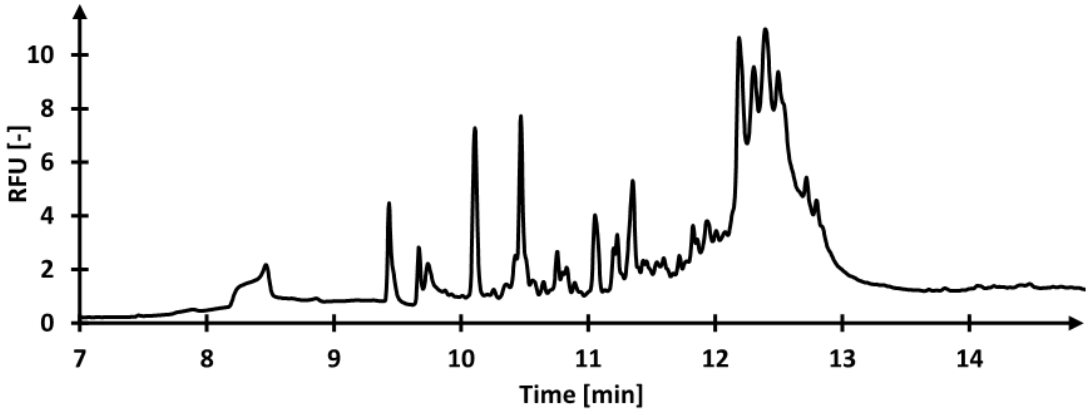


Figure S5: Analysis of the rAAV sample by CGE. Additionally to a cluster of main peaks, also plenty of shorter fragments can be detected. BGE: 2% PVP; 1x TBE; 4 M Urea SYBR-Green II 1:25,000

5 Calculations

qPCR

Amount of viral genomes per mL according to qPCR conducted by manufacturer:

$$\frac{N}{V} = 1.03 \cdot 10^{13}\ \frac{vg}{mL} = 1.03 \cdot 10^{16}\ \frac{vg}{L}$$

The number of particles per volume $\frac{N}{V}$ needs to be divided by the Avogadro constant $N_A = 6.022141\ \cdot 10^{23}\ \frac{1}{mol}$.

$$c = \frac{N}{V \cdot N_A} = \frac{1.03 \cdot 10^{16}\ \frac{1}{L}}{6.022141 \cdot 10^{23}\ \frac{1}{mol}} = 1.7104 \cdot 10^{-8}\ \frac{mol}{L} \approx 17.1\ nM$$

The intact rAAV genome has 2498 nt according to the manufacturer with a molar mass (M) of $770{,}353\ \frac{g}{mol}$. The mass concentration ρ is given by:

$$\rho = c \cdot M = 770{,}353\ \frac{g}{mol} \cdot 1.7104 \cdot 10^{-8}\ \frac{mol}{L} = 1.3176 \cdot 10^{-2}\ \frac{g}{L} \approx 13.2\ \frac{ng}{\mu L}$$

ACE

The ACE measurement provided CPA percentage (CPA%) of 3.1% of the total peak area for the AAV and 8.1% for the plasmid. With the total concentration of the PNA $c_{PNA} = 150\ nM$ this leads to a molar concentration of rAAV/plasmid DNA in the hybridized sample $c_{AAV;sample}$ and $c_{Plasmid;sample}$:

$$c_{AAV;sample} = CPA\%_{AAV} \cdot c_{PNA} = 3.1\% \cdot 150\ nM = 4.65\ nM$$
$$c_{Plasmid;sample} = CPA\%_{Plasmid} \cdot c_{PNA} = 8.1\% \cdot 150\ nM = 12.15\ nM$$

This value only reflects the (+) sense target DNA so a factor $F_{DNA} = 2$ needs to be multiplied. The sample was diluted from 30 µL to 40 µL when the PNA was added after dilution, so a dilution factor $F_{D1} = \frac{4}{3}$ is required. Initially 125 µL of rAAV were used and then eluted after the digest into 30 µL leading to another dilution factor $F_{D2} = \frac{30}{125}$. Applying all these factors leads to an estimate of the actual molar concentration of rAAV DNA c_{AAV}:

$$c_{AAV} = c_{AAV;sample} \cdot F_{DNA} \cdot F_{D1} \cdot F_{D2} = 4.65\ nM \cdot 2 \cdot \frac{4}{3} \cdot \frac{30}{125} = 2.976\ nM \approx 3.0\ nM$$

This value can be transferred with the formulas used above to

$$\rho_{AAV} = c \cdot M = 770{,}353\ \frac{g}{mol} \cdot 2.976 \cdot 10^{-9}\ \frac{mol}{L} = 2.293 \cdot 10^{-3}\ \frac{g}{L} \approx 2.3\ \frac{ng}{\mu L}$$

$$\left(\frac{N}{V}\right)_{AAV} = c_{AAV} \cdot N_A = 2.976 \cdot 10^{-9}\ \frac{mol}{L} \cdot 6.022141 \cdot 10^{23}\ \frac{1}{mol} = 1.792 \cdot 10^{15}\ \frac{1}{L} \approx 1.8 \cdot 10^{12}\ \frac{vg}{mL}$$

For the plasmid $F_{D1} = \frac{4}{3}$ and a different factor $F_{D3} = \frac{30}{100}$ are necessary for comparison with UV, since the intial UV measurement took place in 100 µL:

$$c_{Plasmid} = c_{Plasmid;sample} \cdot F_{D1} \cdot F_{D3} = 12.15\ nM \cdot \frac{4}{3} \cdot \frac{30}{100} = 4.86\ nM \approx 4.9\ nM$$

$$\rho_{Plasmid} = c_{Plasmid} \cdot M_{Plasmid} = 4.86 \cdot 10^{-9}\ \frac{mol}{L} \cdot 3{,}094{,}096\ \frac{g}{mol} = 1.504 \cdot 10^{-2}\ \frac{g}{L} \approx 15.0\ \frac{ng}{\mu L}$$

UV photometer:

This measurement was conducted before the digest in a volume of 100 µL and provided:

$$\rho_{AAV;sample} = 6.1\ \frac{ng}{\mu L} \text{ and } \rho_{Plasmid} = 14.2\ \frac{ng}{\mu L}$$

The molecular weight used here is for dsDNA:

$$M_{AAV} = 1{,}517{,}285\ \frac{g}{mol} \text{ and } M_{Plasmid} = 3{,}094{,}096\ \frac{g}{mol}$$

$$c_{AAV,sample} = \frac{\rho_{AAV,sample}}{M_{AAV}} = \frac{6.1 \cdot 10^{-3}\ \frac{g}{L}}{1{,}517{,}285\ \frac{g}{mol}} = 4.020 \cdot 10^{-9}\ \frac{mol}{L} \approx 4.0\ nM$$

And

$$c_{Plasmid} = \frac{\rho_{Plasmid}}{M_{Plasmid}} = \frac{1.42 \cdot 10^{-2}\ \frac{g}{L}}{3{,}094{,}096\ \frac{g}{mol}} = 4.589 \cdot 10^{-9}\ \frac{mol}{L} \approx 4.6\ nM$$

Initially 125 µL of rAAV were used and extracted into 100 µL so a dilution factor $F_{D1} = \frac{100}{125}$ is needed to obtain ρ_{AAV}. For comparison with qPCR also a DNA factor $F_{DNA} = 2$ is required for the molar value because the above value represents dsDNA.

$$\rho_{AAV} = \rho_{AAV;sample} \cdot F_{D1} = 6.1\ \frac{ng}{\mu L} \cdot \frac{100}{125} = 4.88\ \frac{ng}{\mu L} \approx 4.9\ \frac{ng}{\mu L}$$

$$c_{AAV} = c_{AAV,sample} \cdot F_{D1} \cdot F_{DNA} = 4.020\ nM \cdot \frac{100}{125} \cdot 2 = 6.433\ nM \approx 6.4\ nM$$

$$\left(\frac{N}{V}\right)_{AAV} = c_{AAV} \cdot N_A = 6.433 \cdot 10^{-9}\ \frac{mol}{L} \cdot 6.022141 \cdot 10^{23}\ \frac{1}{mol} = 3.874 \cdot 10^{15}\ \frac{1}{L} \approx 3.9 \cdot 10^{12}\ \frac{vg}{mL}$$

2.4 Electrophoretic characterization of LNP/AAV encapsulated nucleic acids – strengths and weaknesses

Previous studies on NAs from diverse sources have demonstrated their complex nature. No single analytical method has been found to be universally effective in handling all types of NAs holistically. Due to the recognition of certain analytical gaps for longer NAs, a study was launched to demonstrate different analytical methods, highlighting their specific advantages and limitations. The first half of the investigation involved the assessment of mRNA identity and quantity in free state through the utilization of ddPCR and PNA based ACE, as well as the confirmation of pDNA and viral DNA identity and purity via CGE combined with enzymatic digestion of the DNA. The second part of the investigation focused on the measurement of mRNA in a LNP formulation by ACE, CZE, and IP-RP. In addition, a CZE method was described, which allowed for the visualization of the electric charge distribution of complete LNPs.

My involvement in this publication included the acquisition of CE data for the first two chapters of the study, as well as consolidating data from multiple contributors into a comprehensive package. The employed methodologies effectively addressed various CQAs including quantity, identity, purity, and encapsulation efficacy. The different optimal utilization scenarios for every technology were delineated and displayed, that the selection of the analytical approach should be executed on a case-by-case basis.

4TH PROJECT:

Electrophoretic characterization of LNP/AAV encapsulated nucleic acids – strengths and weaknesses

Electrophoresis. 2023 Oct; 44(19-20):1595-606

Received: 5 June 2023 | Revised: 20 July 2023 | Accepted: 30 July 2023

DOI: 10.1002/elps.202300127

ELECTROPHORESIS

RESEARCH ARTICLE

Electrophoretic characterization of LNP/AAV-encapsulated nucleic acids: Strengths and weaknesses

Andrei Hutanu[1,2] | **Miriam López Ferreiro**[1] | **Joost van Haasteren**[1] | **Oliver Höcker**[3] | **Cristina Montealegre**[3] | **Marina Mäser**[3] | **Alex Keresztfalvi**[3] | **Jose Monti**[3] | **Maria Anna Schwarz**[2,3]

[1]Pharma Technical Development, Cell and Gene Therapy Unit, F. Hoffmann-La Roche AG, Basel, Switzerland

[2]Department of Chemistry, University of Basel, Basel, Switzerland

[3]Solvias AG, Kaiseraugst, Switzerland

Correspondence
Maria Anna Schwarz, Solvias AG, Römerpark 2, 4303 Kaiseraugst, Switzerland.
Email: maria.schwarz@unibas.ch

Andrei Hutanu, Department of Chemistry, University of Basel, St. Johanns-Ring 19, 4056 Basel, Switzerland.
Email: a.hutanu@unibas.ch

Color online: See article online to view Figures 3–5 in color.

Abstract

The use of nucleic acids (NAs) has revolutionized medical approaches and ushered in a new era of combating various diseases. Accordingly, there is an increasing demand for accurate identification, localization, quantification, and characterization of NAs encapsulated in nonviral or viral vectors. The vast spectrum of molecular dimensions and intra- and intermolecular interactions presents a formidable obstacle for NA analytical development. Typically, the comprehensive analysis of encapsulated NAs, free NAs, and their spatial distribution poses a challenge that is seldom tackled in its complete complexity. The identification of appropriate physicochemical methodologies for large nonencapsulated or encapsulated NAs is particularly intricate and necessitates an evaluation of the analytical outcomes and their appropriateness in addressing critical quality attributes. In this work, we examine the analytics of nonencapsulated or encapsulated large NAs (>500 nucleotides) utilizing capillary electrophoresis (CE) and liquid chromatography (LC) methodologies such as free zone CE, gel CE, affinity CE, and ion pair high-performance liquid chromatography (HPLC). These methodologies create a complete picture of the NA's critical quality attributes, including quantity, identity, purity, and content ratio.

KEYWORDS
CGE, IP-RP-LC, lipid nanoparticle, peptide nucleic acid, recombinant adeno associated virus

1 | INTRODUCTION

RNA and DNA play an increasingly important role in being used as therapeutics in various pharmaceutical modalities, for example, plasmids [1], (antisense) oligonucleotides (ASO) [2], aptamers [3], gene therapy products [4], and mRNA vaccines [5, 6], which have been a particular focus of attention in the last 3 years. Nucleic acids

Abbreviations: ASO, antisense oligonucleotides; CQA, critical quality attribute; ddPCR, digital droplet PCR; ds, double stranded; GFP, green fluorescent protein; HILIC, hydrophilic interaction chromatography; IP, ion pair; LNP, lipid nanoparticle; NA, nucleic acid; nt, nucleotides; pDNA, plasmid DNA; PNA, peptide nucleic acid; rAAV, recombinant adeno-associated virus; RT, reverse transcriptase; siRNA, small interfering RNA; ss, single stranded.

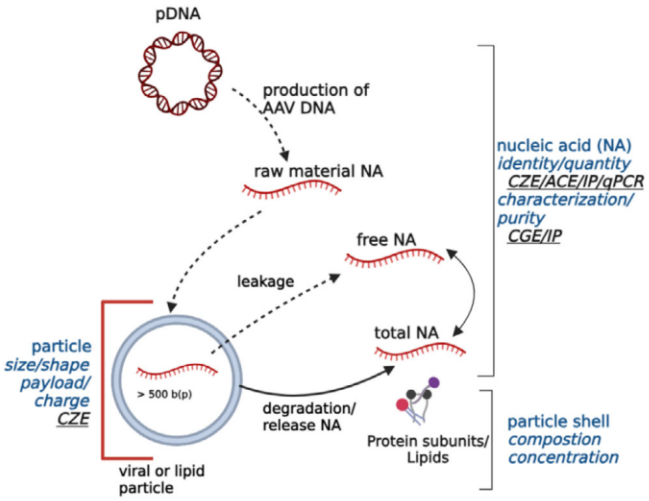


FIGURE 1 Strategic overview of the analytical methods used in this work for characterizing (non-) encapsulated nucleic acid (NA). For more details, see Section 3.

(NAs) are impressive molecules, both from a biochemical and chemical perspective. Due to their intrinsic molecular structure and their remarkable length, they are predestinated to interact in a specific and nonspecific way with many other molecules. Partially, these properties are well described and support the explanation of biochemical reactions where NAs interact with inorganic and organic ions, proteins, peptides, and of course with themselves. Not unusual for large molecules, but a challenge from the analytical perspective, is the fact that NAs change their secondary, tertiary, and quaternary structure significantly depending on the surrounding environment, especially if single-strand NAs are considered [7].

Since the presence of the wrong NA in a formulation is linked to genotoxicity and immunogenic effects, [8] purity and the presence of impurities of NAs are considered critical quality attributes (CQAs) for all NA therapeutics such as lipid nanoparticles (LNPs) [9] or recombinant adeno-associated viruses (rAAVs) [10]. Other NA-related CQAs that can be assessed with physicochemical methods include content/quantity and encapsulation efficiency [9] (see Figure 1). Currently, the testing panel for CQA assessment is dominated by molecular biology techniques; however, it is to be expected that as knowledge on specific modalities increases, physicochemical characterization will be required. In general, for short NAs (e.g., ASO, small interfering [siRNA], or aptamers) analysis, a significantly higher number of techniques are available, whereas for large NAs the choice is more limited. For short single-stranded NA, various options could be useful; ion pair-reversed phase (IP-RP), hydrophilic interaction chromatography (HILIC), and also ion pair-anion exchange chromatography are quite suitable separation technologies that are able to identify, quantify, and track NA-related impurities [11–13]. In most cases, UV detection is applied. A particular advantage of IP-RP and HILIC is that they can also be used with an online coupling to MS, identifying the main components but also traces of impurities at the lower mass range [14]. Capillary electrophoresis (CE), mainly gel CE (CGE), is being widely used for single-stranded NA (ssNA) analysis with applications to the analysis of ASOs or siRNA [15]. However, analysis of ssNA with a significant length (>500 nucleotides [nt]) as mRNAs is more challenging. The NA strand is more instable regarding tertiary structure and moreover enzymatically susceptible. LC and CGE methods show reduced separation performance in the higher mass range [16]. Although CGE shows less instrumental limitations and slightly higher separation efficiency for ssNA than chromatographic methods, it can struggle with robust and reproducible separations [15]. Techniques such as PCR and high-throughput sequencing (HTS) are applied for the study of critical quality attributes of ssNA but also bring their own advantages and disadvantages. PCR remains unrivaled in terms of sensitivity, since, theoretically, it can identify a single target molecule. On the other hand, the possibility of cross-contamination with amplifiable DNA may make this high sensitivity problematic [17]. In addition, only a small portion of the NA gets detected in practice. When just one set of primers is employed, this may lead to inconsistent findings. The holistic approach of HTS is a good choice to verify identity of NA therapeutics with longer sequences, which cannot be targeted with MS, and it can also track down unknown impurities [18]. Nevertheless, HTS remains expensive in routine use due its complexity and the need for specialized staff [17, 19]. Moreover, the library preparation is complicated and can add biases and the steadily changing sequencing instrumentation makes the definition of standards for quality control use difficult. Pre-equilibrium affinity capillary electrophoresis (ACE) seems to be an excellent fit for the ssNA category. It has demonstrated in previous studies that NA with a size of 1000 nt can be specifically detected in the lower nM range [20]. The applicability of ACE for quantification and identification of a large ssNA will be discussed in this paper and compared to other orthogonal techniques such as digital droplet PCR (ddPCR).

Considering double-stranded NAs (dsNAs) the analytical world looks much friendlier than for ssNA due to higher stability in terms of tertiary structure and susceptibility to enzymatical degradation. In general, for the smaller strands, all techniques that are also used for the ssNAs can be considered. The examination of dsNA exceeding 6000 nt may present challenges, as separation techniques are approaching their limits. CGE is well established as a fast, sensitive, and robust alternative to

traditional agarose gel electrophoresis for plasmid DNA purity. However, identification becomes challenging in general for the chromatographic and electrophoretic methods due to the lower resolution. An alternative approach for identification might be charge detection mass spectrometry, which was recently applied to determine the intact molecular mass of AAV-extracted DNA [21]. Enzymatic digestion of larger dsNAs brings us back into a size range that is analytically manageable, especially for the high-resolution techniques such as CGE [22] and IP-RP [23]. This approach could be of great interest for the analytics of adenovirus genomes after capsid degradation. Hereby quantification and identification are feasible. A recent study demonstrated that a highly controlled digestion with RNAse makes this approach also applicable to mRNA [24].

Additionally, a large proportion of NA formulations are not administered in free form to patients but are using some sort of the delivery system. This is to prevent degradation and to increase transduction or lipofection efficiency [1], but makes suitable sample preparation for NA analysis necessary. It is crucial to add that during the test sample preparation, a change in the secondary and tertiary structure is to be expected due to increased temperature, addition of detergents, enzymes, or salts at high concentration, which must be considered in any case. This means that the test sample preparation must also be developed to not only remove the protein or lipid shell and release of NA, but also to ensure that the NA is in a suitable form for analysis. Some efforts have been made for the analysis of mRNA [25–28]. Classical capillary zone electrophoresis (CZE) can offer a significant contribution in the case that the NA is encapsulated into an LNP. CZE can be used to estimate the concentration of free and total RNA, after degradation of the LNP system, but also to observe the entire LNP particle with its RNA filling. Both are very important criteria, which were not accessible for quality control so far, and will be discussed in this study.

The objective of this study is to address some CQAs of NA therapeutics through diverse analytical approaches. Starting from a self-developed strategy based on different analytical methods suitable for the specific type of encapsulated and free NA, the applicability of these methods will be discussed. These methods could serve as a valuable complement to established molecular biology techniques, such as ddPCR, and they cover both RNA and DNA. The initial sections of our study focus on the verification of identity and quantity of mRNA in a free state using ddPCR and ACE as well as plasmid DNA (pDNA) and viral DNA identity and purity by CGE. The next part of the study is centered on the quantification of mRNA present in an LNP formulation and the methods used to determine the total and free NA content, specifically ACE, CZE, and IP-RP. Finally, we will introduce a novel CZE approach that enables the observation of the electrical charge distribution of intact LNPs.

2 | MATERIALS AND METHODS

2.1 | Reagents and test sample material

NaOH (Cat. no. 28244.295) was obtained from VWR Chemicals. HCl (Cat. No. 387800010) was purchased from Acros Organics and 10% SDS solution (Cat. No. 1610416) was obtained from Bio-Rad Laboratories. The following chemicals were obtained from Sigma-Aldrich/Merck KGaA: gradient-grade acetonitrile (Cat. No. 100030), 1 M triethylammonium acetate (TEAA, Cat. No. 625718), ammonium acetate (Cat. No. BCCG2345), ethanol (Cat. No. 100983), TBE buffer 10x (Cat. No. 574795), 0.5 M EDTA (Cat. No. 15575020), formamide (Cat. No. 47671), polyvinylpyrrolidone (PVP; Cat. No. 437190), 10x Tris borate EDTA (TBE) buffer (890 mM Tris, 890 mM boric acid, 20 mM EDTA) (Cat. No. 574795), Tris Base (Cat. No T6066), TritonX-100 (Cat. No. T8787), and urea (Cat. No. U6504). Nuclease-free water (not DEPC-treated) (Cat. No. 4387936) and SYBR Green II RNA Gel Stain 10,000 × concentrated in DMSO (Cat. No. S7564) were purchased from Thermo Fisher Scientific. The QIAquick PCR purification kit (Cat. No. 28104) used for DNA purification was obtained from Qiagen. Eurofins Genomics synthesized all primers. All peptide nucleic acids (PNAs) were obtained from PNABio. The sequences of all nucleic acids used in this study and the sequence of the enhanced green fluorescent protein (EGFP) gene can be found in the Supporting Information. EaeI (Cat. No. R0508) and BtgZI (Cat. No. R0703) nucleases were obtained from New England Biolabs Inc.. Water of HPLC grade was prepared in a Milli-Q-Station (Merck Millipore/Merck KGaA). Solutions were filtered through 0.2 µm membrane filters (Corning).

rAAV production plasmid DNA (Cat. No. VB190926-1395dab) and a corresponding rAAV samples (serotypes AAV2, AAV3, and AAV5) were from VectorBuilder Inc.. CleanCap EGFP mRNA (Cat. No. L-7601, chapter 3.1) was obtained from TriLink Biotechnologies. LNP1 and LNP2 were obtained from different commercial manufacturers. The incorporated mRNA should express EGFP. mRNAs have a theoretical size of approximately 1000 nt for Supplier 1 (LNP1) and >2000 nt for Supplier 2 (LNP2). The exact lipid composition of the LNPs is kept proprietary by the manufacturers. The mRNA (reference, chapter 3.3) from Supplier 1 was also obtained in a pure form.

2.2 | Methods

2.2.1 | One-step ddPCR

The EGFP mRNA was diluted 5×10^8 times from the stock concentration (≈1 mg/mL), resulting in a final concentration of 2×10^{-9} mg/mL. Non-DEPC treated water was used for the dilutions. The reaction mixtures were assembled using a ddPCR One-Step Supermix kit (Bio-Rad; Cat. No. 1864022) with 900 and 250 nM as final concentration for primers and probes, respectively, in a final volume of 24 µL. The PCR plate was sealed and centrifuged at 2000 rcf for 2 min, vortexed at 1400 rpm for 1 min, and centrifuged again at 2000 rcf for 2 min. The samples were transferred to a QX ONE cartridge (Bio-Rad; Cat. No. 12006859), sealed with a PCR plate heat seal foil (Bio-Rad; Cat. No. 12006843), and centrifuged at 1150 rcf for 30 s. The plate was then scanned in the QX ONE Droplet Digital PCR system from BioRad, and the samples were run using the following thermal cycling conditions: 25°C for 3 min, 45°C for 60 min, 95°C for 10 min, 40 cycles of 95°C for 30 s and 60°C for 1 min, followed by a heat treatment at 98°C for 10 min and a final step at 25°C for 1 min before the droplets were read for flourescein signal. The data were analyzed with the QX ONE software 1.2 Standard Edition from Bio-Rad. The concentration per well was obtained in copies per microliter. Only values in the range 50−5000 copies/µL were considered (below 50 copies/µL was considered as the limit of detection and above 5000 copies/µL as a saturated signal). The EGFP mRNA concentration was then calculated taking into account the value obtained from the reading and correcting for the dilution plus the dilution in the PCR plate well (6 µL of sample added in a total volume of 24 µL) and the molecular weight of the mRNA molecule.

2.2.2 | Capillary electrophoresis

Analysis was carried out using a SCIEX PA800 Plus system equipped with a solid-state laser with an excitation wavelength of 488 nm and a 520 nm band pass emission filter (Cat. No. 65–699) from Edmund Optics, a 30 kV power supply, and a temperature controlled auto sampler (±2°C). Data were acquired and analyzed using 32 Karat software 10.3.

2.2.3 | ACE

Separation conditions were as in our previous report [20]. A neutral capillary (Cat. No. 477441) from SCIEX with an I.D. of 50 µm and cut to 20/30 cm was employed. The background electrolyte (BGE) used in all PNA experiments was 6.6x TBE, 3 M urea, and pH was controlled to be 8.3 ± 0.1, and analysis was run at −30 kV.

2.2.4 | CGE

The methodology employed in this study was consistent with that outlined in a prior publication [29]. The gel buffer consisted of 1% (LNP RNA) or 4% (pDNA and rAAV DNA) PVP and 4 M urea in 1x TBE solution with 1:25,000 diluted SYBR Green II. Analysis was performed at −6 kV (pDNA and rAAV DNA) or −3 kV (LNP RNA) in a 20/30 cm fused silica capillaries with 50 µm I.D. (Cat. No. TSP-050375, Polymicro Technologies/Molex LLC).

2.2.5 | CZE

The BGE was 25 mM Tris at pH 8.0 and 1:10,000 diluted (1/100) SYBR Green II. Analysis was performed in a 30/40 cm, 50 µm I.D. fused silica capillary at 30 kV with normal polarity for total LNP analysis and reverse polarity for mRNA direct RNA analysis. Prior to analysis, capillaries were conditioned for 1 min with 0.1 M NaOH (for the normal mode) and with conditioning solution (SCIEX, Cat. No. 713333) for 1 min (for the reversed mode). Sample injection was performed hydrodynamically at 1 psi for 10 s.

2.2.6 | IP-RP

Analyses were performed on an Agilent 1260 Infinity II HPLC, equipped with a quaternary pump, autosampler with flow through needle injector, thermostated column compartment followed by a diode array detector (DAD). Analyses at 70°C were performed using the G7116A column compartment with standard flow pre-heater. The column outlet was directly connected to the detector inlet without post-column temperature control. Signals were acquired at 260 nm. The modules were connected with the shortest possible 0.12 mm I.D. stainless steel capillaries. The nominal volume of DAD flow cell was 1 µL. The gradient delay volume of the system was measured to be 680 µL. System control, data acquisition, and evaluation were performed using Chromeleon 7.2 CDS (Thermo Fisher Scientific). The DNAPac RP column (50 × 2.1 mm, 4 µm, Cat. No. 088924) used in the study has been purchased from Thermo Fisher Scientific. Mobile phase A was water containing 100 mM TEAA (pH 7.0), and mobile phase B was water containing 100 mM TEAA (pH 7.0) mixed with 25 v/v% acetonitrile. The applied gradient was started with an isocratic step of 0% B for 0.6 min, then the following linear increments were set: 4% B at 0.75 min,

39% B to 2.6 min hold till 2.7 min, 51% at 7.5 min, 55% B at 12.5 min, 55.6% B at 23.7 min, 60% B at 30 min, and 100% B at 35 min. For equilibration, it was decreased to 0% B at 36 min and hold for 14 min.

2.3 | Test sample preparation

2.3.1 | Hybridization PNA to mRNA for ACE

PNA stock solutions were 25 μM in formamide. The hybridization conditions were as follows: 15 mM Tris and 2 mM SDS together with 500 nM PNA (approximately threefold excess of PNA to mRNA). When LNP formulations were analyzed during ACE analysis, 1% TritonX was added. This mixture was heated to 95°C for 10 min in a Thermomixer R (Eppendorf SE) and then cooled down slowly for 90 min until the block reached a temperature below 40°C and subsequently analyzed in CE. Calculations for mRNA concentration were performed as described in our previous work [20].

2.3.2 | Digest of plasmid and rAAV DNA

Prior to digestion, 125 μL rAAV genome was purified with the QIAquick PCR purification kit as described in the manufacturer's protocol or our previous report [29] and eluted in 100 μL TE buffer. Afterward, the sample was heated at 95°C for 10 min and cooled down slowly for 90 min in the heat block to allow hybridization. DNA concentration was measured with a NanoDrop 2000 spectral photometer (Cat. No. ND-2000) from Thermo Fisher Scientific. The plasmid stock was diluted in TE to the same molar concentration and then the concentration was checked with the photometer. The EaeI digest was performed in a total volume of 119 μL containing 30 U (6 μL) of EaeI, 12.5 μL digestion buffer (rCutSmart 10x; provided with the nuclease), and 2 μg (virus) or 4 μg (plasmid) of DNA at 37°C for 60 min. Afterward, 30 U (6 μL) of BtgZI was added to a total volume of 125 μL and the samples were incubated at 60°C for 60 min. Purification of the digested DNA was performed again using the QIAquick PCR purification kit.

2.3.3 | LNP test sample solution for CZE

The LNP test sample solution was either directly injected or incubated with TritionX (1%−3%, v/v) for 10 min at different temperatures (25°C, 30°C, 60°C, and 70°C). In the case of the analysis of the intact LNP, 10% v/v Sybr green II (1:100 dilution of SYBR Green II) was added to the test sample.

2.3.4 | LNP mRNA precipitation for IP-RP

Note that 90 μL of the LNP test sample (0.1 mg/mL) or 45 μL of the reference mRNA (1.0 mg/mL) was mixed in a ratio of 9:1 with 5 M ammonium acetate buffer. Afterward, the solution was mixed in a ratio of 1:3 with ice cold ethanol and incubated overnight at −20°C. After the incubation, the mixture was centrifuged at 16,100 *g* at 4°C for 15 min, the supernatant was discarded, and the pellet was washed with 400 μL 70% ice cold ethanol with the same centrifugation conditions. The resulting pellet was dried for 1 h at room temperature under a flow hood. The pellet was resuspended in 1× TBE to a final volume of 50 μL. Injection volumes (HPLC) were set to 2.0 μL for LNP samples and 0.2 μL for mRNA samples.

3 | RESULTS AND DISCUSSION

3.1 | Identity testing and quantification of free mRNA

In this article, the focus was laid on the analysis of NA raw material (see Figure 1). To this end, quantification and unambiguous identification of an green fluorescent protein (GFP) encoding mRNA were performed using ddPCR and an orthogonal pre-equilibrium ACE approach. ddPCR is a form of PCR that involves partitioning the PCR reaction mixture into nanoliter-sized droplets. These droplets contain zero, one, or multiple copies of the target nucleic acid template. Individual droplets can be analyzed, and based on the fluorescence intensity of a specific droplet, it is possible to ascertain if a target molecule was present in the droplet prior to the initiation of thermal cycles. The main benefits of the technique are a very high sensitivity, specificity, good reproducibility, and absolute quantification without the need for a standard curve [30]. However, when RNA is analyzed, reverse transcription (RT) is required to generate complementary DNA. This may add a bias to the test result as this step needs its own primers. A related way to quantify RNA is through the use of a PNA probe that hybridizes with the target molecule and subsequent analysis by CE. The concentration of mRNA in the sample is calculated based on the amount of PNA–mRNA complex that is detected as described previously [20]. For an initial comparison of the techniques, an mRNA sample coding for EGFP with 996 nt and a concentration of 1132 ± 24 μg/mL (determined by UV measurement) was used. The EGFP mRNA concentration obtained through one-step ddPCR was 880 ± 54 μg/mL. In the 1D amplitude graph (Figure S1), a clear separation between the positive and the negative droplets with a high amplitude was observed.

On the other hand, the ACE method obtained a concentration of 1161 ± 6 µg/mL. Our experiments led us to two conclusions. First, the ACE and UV measurement are very similar indicating that the sample contains little to no process-related impurities. Second, the lower value for ddPCR is most likely linked to the RT efficiency. However, the lower value could also arise through a different specificity: the primers used are covering a larger part of the gene than the two PNAs used in ACE (100 vs. 40 nt). Overall, ddPCR is the method of choice when a low sensitivity is crucial, while the PNA hybridization approach is most powerful for RNA samples that are higher in concentration. The main benefit of the ACE approach is that it avoids the RT and amplification steps. Through this direct detection, the test sample preparation is easy to handle and the overall process is more robust. However, for both techniques, accurate quantification is only feasible if no RNA fragments are expected. In order to confirm integrity, it is necessary to employ size-based separation methodologies in parallel, as will be demonstrated subsequently.

3.2 | Identity testing and purity assessment of rAAV DNA

DNA analysis of large molecules can be difficult as the resolution of most techniques decreases with rising nucleotide counts. One potential method to overcome this constraint involves the enzymatic hydrolysis of dsDNA, resulting in the generation of shorter fragments. This technique is widespread as an identity test for plasmids in the laboratory setting and is typically performed via agarose gel electrophoresis. In the case of intricate gene therapy products, enhanced resolution for purity determination and identity data can be achieved by conducting a simultaneous digestion of the production plasmid and the viral DNA. The therapeutic DNA would possess a distinct peak profile, which is juxtaposed against the pDNA profile. While the identity can also be covered by ddPCR or ACE, the purity profile adds significant value here.

In this section, we present this approach for the rAAV production plasmid (containing an EGFP transgene) and the corresponding rAAV therapeutic DNA (Figure 2A). Several batches and serotypes of rAAV were tested, so the aspect of purity assessment could be demonstrated. The digestion process was designed to produce fragments of less than 1500 nt, thereby ensuring that the method's resolution falls within the optimal range. Although rAAVs have a ssDNA genome, there is a mixture of (+) and (−) polarity DNA encapsulated, so dsDNA starts to form as soon as the DNA leaves the capsid [31, 32]. It is noteworthy that sufficient time must be allocated to the hybridization process to ensure almost complete dsDNA formation. The

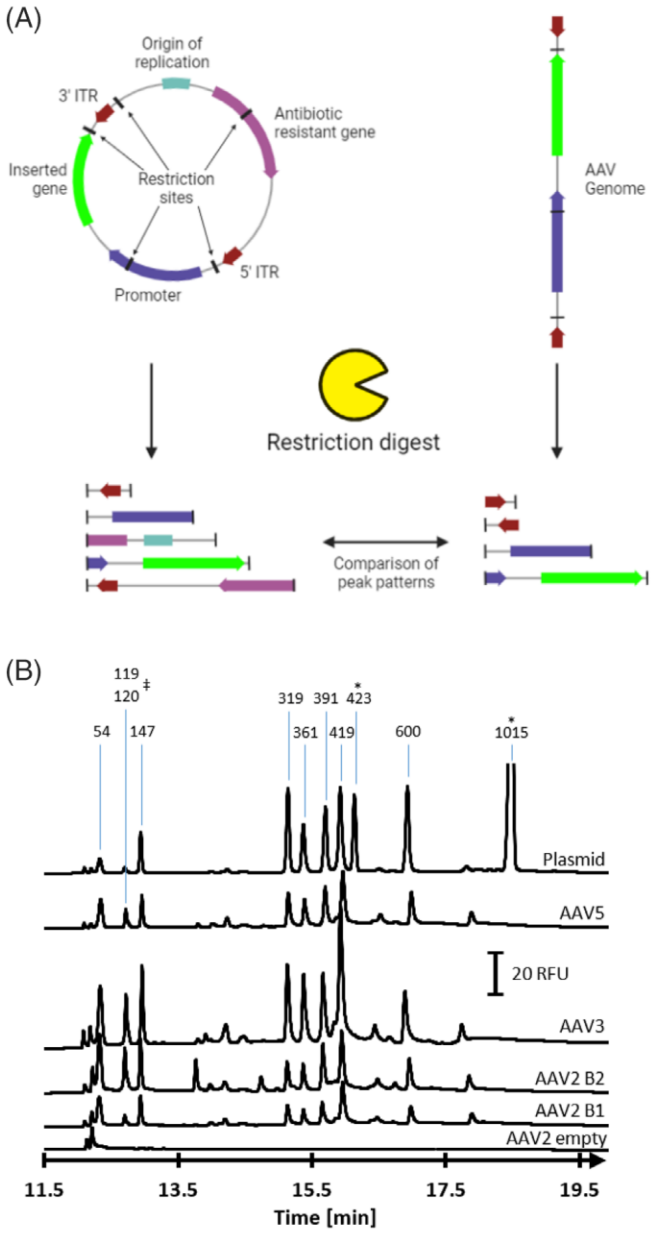


FIGURE 2 **(A)** Simplified scheme of a DNA fingerprinting experiment with the recombinant adeno-associated virus (rAAV) production plasmid with the rAAV genome. It is noteworthy that the plasmid consists of double standard DNA (dsDNA), while the single-standard DNA (ssDNA) of the rAAV needs to be hybridized to dsDNA first before digestion is possible. **(B)** Comparing digested plasmid and rAAV samples reveals batch-specific impurities. rAAV DNA of four different batches that contained the same genome was purified, hybridized, and then digested with *Eae*I and *Btg*ZI restriction nucleases. An empty AAV2 batch that was produced without the vector plasmid was treated the same way as blank. After a second purification step, samples were analyzed with gel CE (CGE). The length of the fragments is indicated on the top. Fragments marked with "*" are only expected in the plasmid sample, while those marked with "‡" are only expected in the rAAV samples. Two different AAV2 batches were measured, namely, AAV2 B1 and B2.

dsDNAs that were obtained during test sample preparation of rAAVs and the corresponding plasmid underwent digestion using identical enzymes, and subsequently, the peak patterns were compared (Figure 2A). In addition, an *in silico* digestion was performed to assign peaks. Due to variations between pDNA and therapeutic DNA, certain peaks were anticipated to solely belong to the plasmid (denoted by asterisk) and others exclusively to the therapeutic DNA (denoted by ‡; Figure 2B). Data from the *in silico* digestion implied that there should be seven main signals (54, 147, 319, 361, 391, 419, and 600 bp) present in pDNA and rAAV samples. The actual electropherograms of all rAAV batches in comparison to the plasmid clearly demonstrated that these signals are present, thereby confirming the DNA identity.

In addition, comparison among batches and the plasmid digests exhibited unexpected minor peaks in between 147 and 319 bp as well as 423 and 1019 bp (Figure 2B). Some of them are present in all performed runs. These are plausibly artifacts of the methodology owing to aberrant enzymatic activity and could not be assigned to any plausible miscleavage. However, these signals may represent impurities that are present in the production plasmids, and subsequently get transferred into the rAAV capsid during manufacturing. Peaks that are specific to a particular batch are likely to be caused by variations in the manufacturing process and therefore require special attention. In the presented example, four peaks of the rAAV exhibited a high degree of proximity during separation, with their respective length spanning from 319 to 419 nt. The formation of the shoulder peaks, that is, at the 419 nt peak of AAV3, suggests the presence of additional unexpected DNA fragments, which could represent impurities and are "hidden" behind these signals. In case characterization of these signals is the primary objective, the electropherograms need to be compared to undigested samples and with samples that were digested by different enzyme sets. Through the high specificity of the used enzymes, signals detected prior to digestion are suspected to be impurities with an unknown sequence, and their mobility should remain unchanged after the digestion process. If novel signals appear after digestion and the presumed contaminant disappears, it is probable that this NA is a constituent of the intended DNA. However, further confirmation is required by employing alternative sets of enzymes. The digestion approach is not restricted to rAAV alone, but could be extended to any dsDNA therapeutics, such as adenoviruses or herpesviruses. Furthermore, it could potentially exhibit transferability toward ssNA therapeutics through the utilization of a Cas9 nuclease that cleaves ssNA [19]. In this scenario, the production of detectable short fragments would require the utilization of multiple single guide RNAs. Nevertheless, it seems that the commercial unavailability of these altered Cas9 versions would pose a significant challenge to their routine testing and application.

3.3 | Characterization and quantification of LNP mRNA

The ensuing discussion centers on the quantification and characterization of mRNA formulated into an LNP-containing solution. The essential question hereby is: What is the concentration of the RNA both inside and outside of the LNP, and in which form is the RNA present? In classical CZE, the separation of NA size variants is hindered by their constant hydrodynamic radius to charge ratio. However, CZE can detect the total comigrating RNA variants present in a formulation in their free form by directly injecting the test sample without any additional treatment. For this, the analysis was performed at negative polarity with SYBR Green II present in the BGE. Under these conditions, RNA encapsulated in LNPs will not be detected because the overall charge of the LNPs inhibits migration to the anode. Quantification of free RNA would be certainly possible by utilizing an RNA reference material with a predetermined mass concentration. An identical methodology can be employed to measure the total mRNA content when the LNP system is disrupted with a suitable surfactant under appropriate conditions prior to analysis. Through combination of both approaches, the ratio of free to total mRNA can be determined, which is a CQA due to its substantial influence on the product's effectiveness. The analysis of free mRNA is compared to a test sample containing the mRNA encapsulated into an LNP system, which was released through TritonX treatment at different temperatures, as shown in Figure 3.

As the increasing temperature and the presence of the nonionic surfactant lead to disruption of the LNP structure, the signal intensity increases. The data suggest that the amount of total RNA present in the LNP system is released at 60°C. For the chosen test sample, it was possible to determine the signal intensity ratio between free and total RNA to be about seven times. It is noteworthy that the analytical outcome of this CZE assay does not provide any inference about the state of the RNA, such as the progress advancement of degradation or any other chemical modifications.

In order to focus on potential degradation products, a CGE analysis was performed based on the same approach: untreated test sample (displaying the mRNA that is localized outside of the LNPs, or free RNA) was compared with the result of the test sample in which the entire RNA is in a state of free dissolution (Figure 4A). Compared to

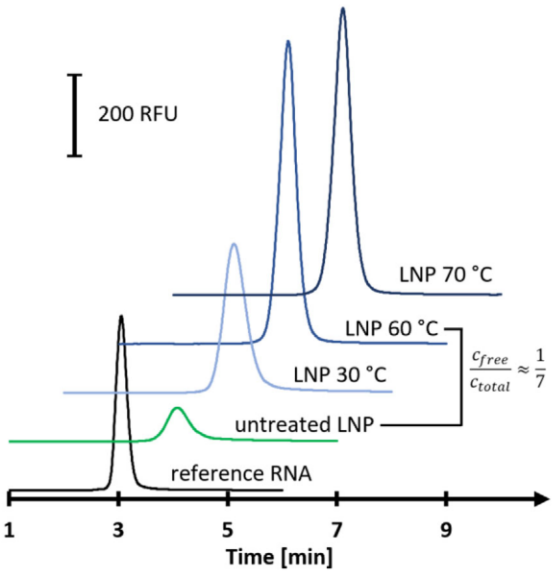


FIGURE 3 Treatment of lipid nanoparticle (LNP) formulations with nonionic surfactants at different temperatures. Overlay of capillary zone electrophoresis (CZE) electropherograms of reference material: reference mRNA (black), untreated LNP test sample, displaying the mRNA that is localized outside of the LNPs or free mRNA (green) and released mRNA after treatment with TritonX at given temperatures, total mRNA (blue).

CZE, not only signal intensities but also the separation profile of RNA size variants can offer valuable insights. This observation is potentially significant since the degradation process within the LNP may exhibit distinct characteristics compared to degradation outside the nanoparticle's protective environment. Accurately quantifying the variants can be challenging, particularly in case of complex profiles or significant degradation (see Figure 4A, insert). The underlying cause for this is that the signal intensity is proportional to the mRNA size and intercalating dyes exhibit a specific affinity toward different nucleobases [33]. Furthermore, the laser-induced fluorescence (LIF) response can vary considerably within an analytical CE sequence.

In addition, beyond the utilization of CGE and CZE, the investigations were expanded to ACE and IP-RP for quantifying RNA in LNP formulations that underwent either TritonX treatment or RNA precipitation. ACE was shown in Section 3.1 to be well suited for mRNA in free solution, but it is noteworthy that there may be certain challenges associated with utilizing PNAs for the purpose of RNA quantification in LNP formulations. The LNPs might interfere with the hybridization of the PNA probe to the RNA molecule, or alternatively, the LNPs could impede the separation and subsequent detection of the PNA–RNA complex. Nevertheless, these challenges should be reduced through the disintegration of the LNPs with TritonX. The mRNA level of LNP1 was quantified in a solution containing 1% TritonX by utilizing two PNAs that bind to GFP. The observed value of 95 µg/mL was in close proximity to the manufacturer's specified concentration of 100 µg/mL, indicating the efficacy of the ACE methodology in detecting and measuring mRNA levels in complex formulations.

IP-RP was also tested for the same set of LNP samples in order to understand the capabilities of HPLC to quantify and characterize LNP mRNA. Due to strong disturbance of TritonX in the HPLC system, both LNPs were prepared by RNA precipitation. Quantification in the presence of lipids through precipitation is generally challenging due to the highly variable recovery, which can be below 50% in certain cases. These values were obtained utilizing a UV spectrophotometer and CGE (data not shown). The strong interaction between lipids and NA could be the reason for the reduced recovery in this scenario. To ensure precise quantification, additional investigations are required regarding the preparation and handling during precipitation. This would be intriguing, as quantification using a UV detector in HPLC can be deemed more straightforward compared to quantification using CE-LIF with fluorescent dyes. Comparing the profiles of the total RNA for LNP1 in Figure 4A,B (green traces), it seems that with IP-RP less size variants are revealed. Differences in sample preparation could be a possible cause for this observation, because the separation efficiency in this mass range should be comparable to CGE (data not shown).

The analytical results clearly showed that both CGE and the IP-RP require further development and investigations, especially for the test sample preparation. It is imperative to enhance our comprehension of the higher order structures of ssNA during test sample preparation and separation processes. This is decisive for effectively controlling the secondary and tertiary structure during analysis, leading to improved interpretation of analytical results. The utilization of techniques such as circular dichroism would be a fitting approach for this purpose. Investigating the higher order structure of ssNA in the presence of ion pair reagents, detergents such as TritonX or denaturants like formamide, would be of great use for the interpretation of size-based techniques.

3.4 | Characterization of intact LNP

Apart from addressing the localization of mRNA in LNP formulations, the nanoparticles' size and charge distribution are of significant interest. In this section, a conventional CZE is employed for the examination of complete LNP formulations that comprise RNA. The process involves the use of a fluorescent dye to stain the mRNA, which could either be located within the particle or linked directly to the lipids. To this end, SYBR Green II was

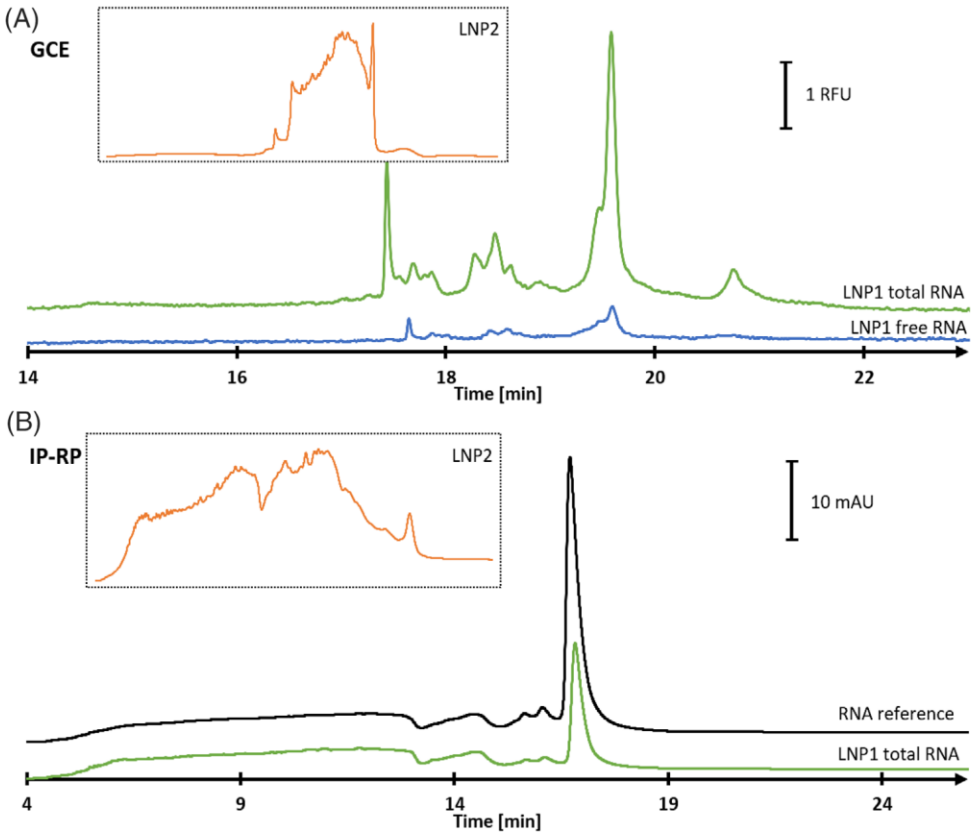


FIGURE 4 Comparison of gel CE (CGE) (A) and ion pair-reversed phase (IP-RP) (B) for the assessment of size distribution of mRNA present in two different lipid nanoparticle (LNP) formulations. (A) Free RNA in LNP1 (blue) and total RNA in LNP1 after treatment with triton (green). Insert: Size distribution pattern for RNA in LNP2 after stress. (B) Total RNA in LNP1 after precipitation (green) and RNA reference without LNP (black). Insert: Size distribution pattern for RNA in LNP2 after stress.

pre-incubated with the samples and also incorporated into the BGE. A recent study from Brader and colleagues found that cationic dyes can permeate the LNP without disturbing its morphology [28]. As opposed to utilizing nonspecific interacting reagents to stain the lipid particle, a direct RNA binding dye offers a significant advantage as the signal intensity is directly proportional to the concentration of the enclosed RNA. Additionally, it enables the identification of various RNA-containing species at their anticipated electrophoretic migration velocity. The determination of the ionic mobility provides insight into the overall charge and surface charge of the nanoparticle, ultimately yielding information on the zeta potential. Due to the formation of an electrochemical double layer caused by the surface charge, the LNP particle acquires a specific charge when it moves in an electrolyte, which corresponds to charge located at the slipping plane. The charge to hydrodynamic radius distribution is anticipated to be individual for each LNP system. Predicting of the distribution is not possible, but it is highly probable that it correlates with the changes in morphology during stress studies of the LNP system.

In our study, we have chosen to examine two distinct LNP mRNA systems that vary in RNA size, concentration, and lipid composition (LNP1 and LNP2). We selected a simple BGE (25 mM Tris pH 8.0), which is expected to have no impact on the LNPs during the analysis. Upon initial observation, Figure 5 displays notable differences in migration time, profile, shape, and signal intensity between the electropherograms of the different LNPs, which can be attributed to the inherent characteristics of the LNP–RNA systems. At the selected pH, the main signal associated with the LNPs exhibits a slightly negative net charge (migrating behind the electroosmotic flow [EOF]). However, LNP2 displays a more negative ionic mobility/charge and a sharper peak shape than LNP1, indicating a lower charge to hydrodynamic radius distribution. Together with the primary signal in the normal polarity mode, signals in proximity to the EOF with reduced negative mobility were observed in both samples (Figure 5). The species in question may be associated with LNP aggregates that exhibit RNA-binding properties, as suggested by repeatedly nonreproducible profiles in this region. The signals detected in this normal polarity mode disappeared upon treatment of the test sample with TritonX (data not shown), indicating disruption of the LNP. Conversely, the signal intensity of mRNA in the reverse polarity mode is

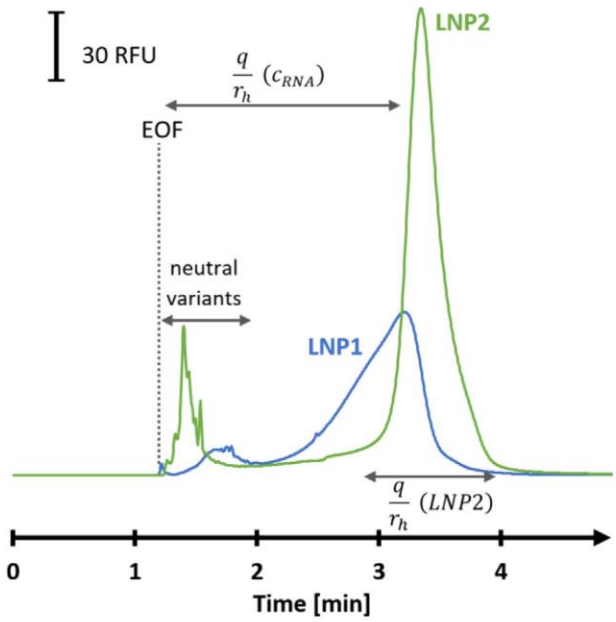


FIGURE 5 Capillary zone electrophoresis (CZE) analysis of two intact mRNA lipid nanoparticle (LNP) formulations reveals the charge to hydrodynamic radius (q/r_h) distribution of RNA and LNPs. The dashed line shows the approximate migration time of the electroosmotic flow (EOF) (determined in separate UV measurement).

amplified in accordance with the total RNA content (see Section 3.3).

In order to estimate the charge and charge distribution, it is vital to initially compute the ionic mobility. To achieve this, the experimental determination of the EOF is required. Neutral fluorescent molecules, such as coumarin dyes, are good candidates to accomplish this task. In the LNP2 system, an observed mobility of -3.5×10^{-4} cm^2/Vs was recorded, while in the LNP1 system, a mobility of -3.4×10^{-4} cm^2/Vs was found. Thus, the range of ionic mobility within which the primary signal is observed can be computed, along with that of any other variants present in the electropherogram. For example, the range for LNP2 falls within -3.0×10^{-4} cm^2/Vs and -4.1×10^{-4} cm^2/Vs. Assuming a constant hydrodynamic radius, the charge of a particle exhibits a direct proportionality to its ionic mobility. By determining the hydrodynamic radius through techniques such as dynamic light scattering, it becomes possible to assess the average charge of the particle.

The precise quantification of concentration may present challenges due to the complex calibration process and the influence of the lipid shell's shielding effects. However, it is possible to effectively observe relative changes by comparing the system to a reference LNP-containing system.

To achieve reproducible results, it is crucial to carefully select the preparation method for the test sample. First, the establishment of a chemical equilibrium in the distribution of dye between the various phases, both external and internal to the particle. In comparison, the significance of incubation temperature and duration is lower, although these variables must still be evaluated in a final stage. These parameters have to be carefully balanced around room temperature, so that the particle is still intact before measurement. Starting from a predetermined concentration of dye, which was approximately tenfold greater in the test sample solution than in the BGE, the dye concentration was altered by ±20%, and the resultant signal intensity was assessed. If the intensity remained constant, it was inferred that saturation had occurred and the chemical equilibrium has been reached.

Assuming optimal conditions are chosen, CZE proved to be valuable in the individual characterization of diverse LNP systems. This includes the determination of the charge /hydrodynamic radius and corresponding distribution of the LNP, and the relative concentration of encapsulated RNA. The comprehensive characterization of intact LNPs could provide significant value for identification and quantification purposes, as well as for stability, comparability, and forced degradation studies. This method/approach could be a valuable extension of the pool of analytical methods, which are helpful for the characterization of nano particles in general. However, in order to establish the method with a significant, scientific relevant value, various complementary methods such as zeta potential or AF4 measurements should be implemented in a systematic way. Moreover, of utmost significance for these fundamental investigations is the commercial availability of diverse LNP systems, with well-defined compositions and concentrations of the manifold constituents. This would greatly aid to scrutinize LNPs stability attributes, storage prerequisites, or conditions for incorporation of specific and nonspecific dyes.

4 | CONCLUDING REMARKS

The present investigation demonstrates that the primary aim of analytical development for NA therapeutics is to integrate the distinct advantages of various techniques to attain a comprehensive and precise analytical result. CZE, ACE, and ddPCR are all viable options for quantifying NA. Nevertheless, all of them encounter challenges as well. The use of designated probes/primers in ACE and ddPCR may result in neglecting fragmentation, whereas in CZE, comigration of all NA species could lead to an overestimation of the drug substance. While CZE's main advantage is its simplicity and speed, ACE and particularly ddPCR offer the benefit of superior specificity and sensitivity. Nonetheless, CZE is valuable for assessing

the ratio of nonencapsuled to encapsulated NAs in LNP formulations.

ACE and ddPCR are essential methodologies for identity testing, and the advantages and disadvantages mentioned above are also applicable for this objective. For dsDNA, an alternative is CGE with previous enzymatic digestion. The primary advantage of this approach is its ability to provide data on both purity and identity, resulting in a comprehensive picture. The use of IP-RP for this purpose is certainly possible; however, it was not investigated in detail in the present study. In general, CGE and IP-RP are the most potent options for assessing overall purity, including ssNA. Nonetheless, quantification poses a challenge due to the fact that each size variant exhibits a distinct UV/LIF response that is contingent upon the nucleotide counts of each fragment. CGE is expected to contribute significantly to the characterization of encapsulated NA and their stability-indicating properties. Equally intriguing is the examination of NA inside and outside the vector with CGE.

CZE analysis of intact LNPs that are permeable for dyes is anticipated to address a significant knowledge gap. For the first time, it was shown that by specific staining of the RNA present in the nanoparticle, both the ionic mobility and its distribution can be directly determined, with the signal intensity being directly proportional to the RNA concentration. This approach makes the technique attractive for the study of stability indicating features, since not only the ionic mobility and thus charge is accessible, but also unexpected new species can be identified. Fundamental investigations may be carried out to examine the physicochemical characteristics of an LNP in relation to its surrounding environment, such as the ionic strength, pH, or formulation buffer composition.

ACKNOWLEDGMENTS

The authors would like Dr. Bernd Moritz for his support reviewing this publication. The authors acknowledge the web tool BioRender.com, which was used to create Figures 1 and 2A.

Open access funding provided by Universitat Basel.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Andrei Hutanu https://orcid.org/0000-0003-0174-5250

REFERENCES

1. Scholz C, Wagner E. Therapeutic plasmid DNA versus siRNA delivery: common and different tasks for synthetic carriers. J Control Release. 2012;161:554–65.
2. Bennett CF. Therapeutic antisense oligonucleotides are coming of age. Annual Rev Med. 2019;70:307–21.
3. Adachi T, Nakamura Y. Aptamers: a review of their chemical properties and modifications for therapeutic application. Molecules. 2019;24:4229.
4. Dunbar CE, High KA, Joung JK, Kohn DB, Ozawa K, Sadelain M. Gene therapy comes of age. Science. 2018;359:eaan4672.
5. Verbeke R, Lentacker I, De Smedt SC, Dewitte H. The dawn of mRNA vaccines: the COVID-19 case. J Control Release. 2021;333:511–20.
6. Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. Nat Rev Drug Discov. 2018;17:261–79.
7. Zhang J, Fei Y, Sun L, Zhang QC. Advances and opportunities in RNA structure experimental determination and computational modeling. Nat Methods. 2022;19(10):1193–207.
8. Davé UP, Cornetta K. AAV joins the rank of genotoxic vectors. Mol Ther. 2021;29:418–9.
9. Blenke EO, Örnskov E, Schöneich C, Nilsson G, Volkin DB, Mastrobattista E, et al. The storage and in-use stability of mRNA vaccines and therapeutics: not a cold case. J Pharm Sci. 2023;112(2):386–403.
10. Wright JF. Quality control testing, characterization and critical quality attributes of adeno-associated virus vectors used for hum gene therapy. Biotechnol J. 2021;16:2000022.
11. Goyon A, Yehl P, Zhang K. Characterization of therapeutic oligonucleotides by liquid chromatography. J Pharm Biomed Anal. 2020;182:113105.
12. Fekete S, Aebischer MK, Imiołek M, Graf T, Ruppert R, Lauber M, et al. Chromatographic strategies for the analytical characterization of adeno-associated virus vector-based gene therapy products. TrAC Trends Anal Chem. 2023;164:117088.
13. Fekete S, Doneanu C, Addepalli B, Gaye M, Nguyen J, Alden B, et al. Challenges and emerging trends in liquid chromatography-based analyses of mRNA pharmaceuticals. J Pharm Biomed Anal. 2023;224:115174.
14. Goyon A, Zhang K. Characterization of antisense oligonucleotide impurities by ion-pairing reversed-phase and anion exchange chromatography coupled to hydrophilic interaction liquid chromatography/mass spectrometry using a versatile two-dimensional liquid chromatography setup. Anal Chem. 2020;92:5944–51.
15. Wei B, Goyon A, Zhang K. Analysis of therapeutic nucleic acids by capillary electrophoresis. J Pharm Biomed Anal. 2022;219:114928.
16. Georgopoulos D, Leibowitz MJ. Use of high-performance liquid chromatographic fractionation of large RNA molecules in the assay of group I intron ribozyme activity. J Chromatogr A. 2000;868:109–14.
17. Baldwin J, Piplani S, Sakala IG, Honda-Okubo Y, Li L, Petrovsky N. Rapid development of analytical methods for evaluating pandemic vaccines: a COVID-19 perspective. Bioanalysis. 2021;13:1805–26.
18. Guerin K, Rego M, Bourges D, Ersing I, Haery L., Harten DeMaio K, et al. A novel next-generation sequencing and analysis platform to assess the identity of recombinant

adeno-associated viral preparations from viral DNA extracts. Hum Gene Ther. 2020;31:664–78.
19. Morris C, Lee YS, Yoon S. Adventitious agent detection methods in bio-pharmaceutical applications with a focus on viruses, bacteria, and mycoplasma. Curr Opin Biotechnol. 2021;71:105–14.
20. Hutanu A, Signori C, Moritz B, Gregoritza M, Rohde A, Schwarz MA. Using peptide nucleic acid hybridization probes for qualitative and quantitative analysis of nucleic acid therapeutics by capillary electrophoresis. Anal Chem. 2023;95(11):4914–22.
21. Barnes LF, Draper BE, Kurian J, Chen Y-T, Shapkina T, Powers TW, et al. Analysis of AAV-extracted DNA by charge detection mass spectrometry reveals genome truncations. Anal Chem. 2023;95:4310–6.
22. Wysoczynski CL, Roemer SC, Dostal V, Barkley RM, Churchill ME, Malarkey CS. Reversed-phase ion-pair liquid chromatography method for purification of duplex DNA with single base pair resolution. Nucleic Acids Res. 2013;41:e194.
23. Huber CG, Oefner PJ, Bonn GK. Rapid and accurate sizing of DNA fragments by ion-pair chromatography on alkylated nonporous poly (styrene-divinylbenzene) particles. Anal Chem. 1995;67:578–85.
24. Vanhinsbergh CJ, Criscuolo A, Sutton JN, Murphy K, Williamson AJ, Cook K, et al. Characterization and sequence mapping of large RNA and mRNA therapeutics using mass spectrometry. Anal Chem. 2022;94:7339–49.
25. Malburet C, Leclercq L, Cotte J-F, Thiebaud J, Bazin E, Garinot M, et al. Size and charge characterization of lipid nanoparticles for mRNA vaccines. Anal Chem. 2022;94:4677–85.
26. Loughney JW, Minsker K, Ha S, Rustandi RR. Development of an imaged capillary isoelectric focusing method for characterizing the surface charge of mRNA lipid nanoparticle vaccines. Electrophoresis. 2019;40:2602–9.
27. Krebs F, Burger U, Dörks S, Kramer M, Wätzig H. Two quality and stability indicating imaged CIEF methods for mRNA vaccines. Electrophoresis. 2022;43:1971–83.
28. Brader ML, Williams SJ, Banks JM, Hui WH, Zhou ZH, Jin L. Encapsulation state of messenger RNA inside lipid nanoparticles. Biophys J. 2021;120:2766–70.
29. Hutanu A, Boelsterli D, Schmidli C, Montealegre C., Dang Thai MHN, Bobaly B, et al. Stronger together: analytical techniques for recombinant adeno associated virus. Electrophoresis. 2022;43:1107–17.
30. Kojabad AA, Farzanehpour M, Galeh HEG, Dorostkar R, Jafarpour A, Bolandian M, et al. Droplet digital PCR of viral DNA/RNA, current progress, challenges, and future perspectives. J Med Virol. 2021;93:4182–97.
31. Gonçalves MA. Adeno-associated virus: from defective virus to effective vector. Virol J. 2005;2:43.
32. Mayor HD, Torikai K, Melnick JL, Mandel M. Plus and minus single-stranded DNA separately encapsidated in adeno-associated satellite virions. Science. 1969;166:1280–2.
33. Zhang B, Jiang J, Yuan Y, Guan Y. Influence of nucleotide-biased fluorescence emissions of SYBR green II on the result consistence of rolling circle amplification. Chem Res Chin Univ. 2019;35:1119–23.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hutanu A, Ferreiro ML, van Haasteren J, Höcker O, Montealegre C, Mäser M, et al. Electrophoretic characterization of LNP/AAV-encapsulated nucleic acids: Strengths and weaknesses. Electrophoresis. 2023;1–12. https://doi.org/10.1002/elps.202300127

SUPPORTING INFORMATION

Electrophoretic characterization of LNP/AAV encapsulated nucleic acids – strengths and weaknesses

Electrophoresis. 2023 Aug; preprint.

The complete sequence of the EGFP gene; green markup was targeted by PNAs; yellow markup by ddPCR.

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG
CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCAC
CACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTAC
CCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG
CTGAAGGGCATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTAC
CAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGA
GCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCA
TGGACGAGCTGTACAAGTAA

Primers and probes. All sequences from 5’ to 3’ end. The following primers and probes were used for ddPCR to detect the EGFP mRNA sequence : forward primer CGCACCATCTTCTTCAAGGA, reverse primer CTTGAAGTCGATGCCCTTCA, probe ACCCGCGCCGAGGTGAAGTTCGA; the 5’ end of the probe was labeled with the FAM fluorophore and the fluorescence was quenched with an internal ZEN quencher and 3’ Black Hole Quencher-1.

PNA probes for ACE were labeled with Oregon green followed by two spacers. Two Glu residues were added for faster CZE mobility and better solubility.

OG488–OO–TCTCGTTGGGGTCTTTGCTC-Glu-Glu;

OG488–OO–TTCTGCTTGTCGGCCATGAT-Glu-Glu

Supplementary Figures

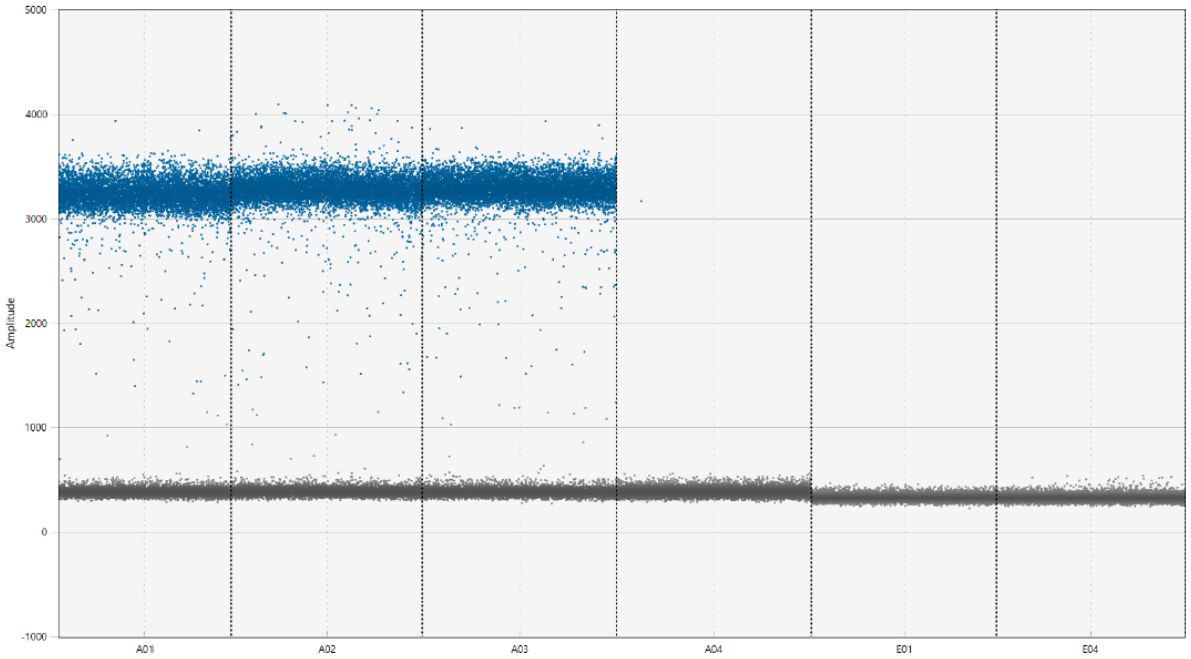


Figure S1: 1D amplitude graph obtained from QX ONE Software: Wells A01, A02, and A03 show the population of both negative and positive droplets in the FAM channel (experiment was performed as triplicate). The well A04 is the non-templacte control, in which no sample was added but the other conditions were maintained. WellE01 and E04 are the equivalent to A01 and A04 but no reverse transcriptase was added and therefore no retro-transcription was possible to obtain DNA for PCR amplification (negative control).

3. Conclusion

This thesis aimed to evaluate the utility of CE in analyzing critical quality attributes of emerging therapeutic agents for their use as innovative drugs, such as peptides, plasmids, mRNA/lipid nano particle formulations, and recombinant adeno-associated virus (rAAV). The main goal was to improve CE's applicability for these modalities by developing novel approaches for these complex therapeutics. The dissertation encompassed four projects that have been presented and discussed in a collection of five articles published in peer-reviewed analytical chemistry journals. The first chapter explored the potential of affinity CE to measure the extent and specific sites of methionine oxidation in peptides and monoclonal antibodies. Thus, it aimed to offer an alternative approach to mass spectrometry. The subsequent chapter presented a comprehensive direct comparison of the current analytical techniques employed in the examination of recombinant adeno-associated viruses (rAAV), encompassing TEM, AUC, CGE, CZE, and IP-RP-LC. In the third chapter, a technique was outlined for the quantification and identification of nucleic acid therapeutics using fluorescent PNA hybridization probes in CE. The concluding section of this dissertation examined different analytical techniques used to analyze larger oligonucleotides (>500 nt) incorporated in diverse delivery systems, emphasizing the advantages and limitations of each method. The methods employed in this study encompassed the utilization of PNA-based CE as described in the previous chapter, ddPCR, CGE with the incorporation of nuclease digestion, IP-RP-LC, and CZE. These techniques were collectively employed to thoroughly analyze NAs and assess their quality attributes as biologics. In conclusion, this dissertation provided a comprehensive analysis of various techniques and methods utilized in the study of biopharmaceutical molecules and molecular systems, emphasizing their interconnectedness. The examination of minor alterations within complex molecules and the characterization of large molecules collectively outlined the broad applicability of CE based methods for new modalities. The results presented in the scientific publications serve as a robust foundation for future research and development for analytical laboratories in the biopharmaceutical sector.

4. Tools and References

3.1 Tools

- DeepL Translate ; Deepl SE: https://www.deepl.com/translator
  - Translation of text passages; entire document.

- QuillBot ; QuillBot (Course Hero), LLC : https://quillbot.com
  - Paraphrasing of text passages; Used in the introduction section.

- Biorender; Science Suite. Inc. : https://www.biorender.com
  - Generating images: Figure 1-8 and in the publications as cited.

- Microsoft Bing Image Creator (DALL-E); Microsoft Corporation: https://www.bing.com/images/create?form=FLPGEN
  - Generation of the cover image, slightly modified.
  - Used prompt: a chromatogram electropherogram that contains a range of mountains with the matterhorn, digital art, book cover, scientific, minimalist, HPLC, colorful on a royal blue background, 4k

3.2 References

[1] Yeh M-K, Chen Y-C. Biopharmaceuticals. BoD–Books on Demand; 2018.
[2] Ho RJ. Biotechnology and biopharmaceuticals: transforming proteins and genes into drugs. John Wiley & Sons; 2013.
[3] Walsh G, Murphy B. Biopharmaceuticals, an industrial perspective. Springer Science & Business Media; 2013.
[4] Szkodny AC, Lee KH. Biopharmaceutical manufacturing: Historical perspectives and future directions. Annual Review of Chemical and Biomolecular Engineering. 2022;13:141-65.
[5] Walsh G, Walsh E. Biopharmaceutical benchmarks 2022. Nature Biotechnology. 2022:1-39.
[6] Weiner GJ. Building better monoclonal antibody-based therapeutics. Nature Reviews Cancer. 2015;15:361-70.
[7] Jiang Z, Dalby PA. Challenges in scaling up AAV-based gene therapy manufacturing. Trends in Biotechnology. 2023.
[8] Bhusnure O, Gholve S, Suryawanshi R, Sugave R, Sangshetti J. Life cycle assessment (LCA) approach to analytical method development: a review. World J Pharmacy Pharm Sci. 2015;4:933-63.
[9] Tremblay G. In: Future Science; 2017. p. 233-4.
[10] Thakur A, Tan Z, Kameyama T, El-Khateeb E, Nagpal S, Malone S*, et al.* Bioanalytical strategies in drug discovery and development. Drug Metabolism Reviews. 2021;53:434-58.
[11] EMA. Guideline on development, production, characterisation and specification for monoclonal antibodies and related products. 2016.

[12] ICH. ICH Topic Q 6 B Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products. 1999.
[13] Geigert J. Quality Attributes of a Biopharmaceutical. The Challenge of CMC Regulatory Compliance for Biopharmaceuticals. 2019:311-29.
[14] Bhandari V, Piccarreta M, Moliner-Szapary P, Rosanna L. Now is the time to design pharma supply chains that can overcome risk today and thrive in the future. 2022.
[15] Ahmadi M, Bryson CJ, Cloake EA, Welch K, Filipe V, Romeijn S*, et al.* Small amounts of sub-visible aggregates enhance the immunogenic potential of monoclonal antibody therapeutics. Pharmaceutical research. 2015;32:1383-94.
[16] Rosenberg AS. Effects of protein aggregates: an immunologic perspective. The AAPS journal. 2006;8:E501-E7.
[17] Alt N, Zhang TY, Motchnik P, Taticek R, Quarmby V, Schlothauer T*, et al.* Determination of critical quality attributes for monoclonal antibodies using quality by design principles. Biologicals. 2016;44:291-305.
[18] Flatman S, Alam I, Gerard J, Mussa N. Process analytics for purification of monoclonal antibodies. Journal of Chromatography B. 2007;848:79-87.
[19] Bhargava AC, Mains K, Siu A, Gu J, Zarzar J, Yi L*, et al.* High-throughput, fluorescence-based esterase activity assay for assessing polysorbate degradation risk during biopharmaceutical development. Pharmaceutical Research. 2021;38:397-413.
[20] Jenkins N, Murphy L, Tyther R. Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. Molecular biotechnology. 2008;39:113-8.
[21] Gao X, Ji JA, Veeravalli K, Wang YJ, Zhang T, Mcgreevy W*, et al.* Effect of individual Fc methionine oxidation on FcRn binding: Met252 oxidation impairs FcRn binding more profoundly than Met428 oxidation. Journal of pharmaceutical sciences. 2015;104:368-77.
[22] Alam ME, Barnett GV, Slaney TR, Starr CG, Das TK, Tessier PM. Deamidation can compromise antibody colloidal stability and enhance aggregation in a pH-dependent manner. Molecular pharmaceutics. 2019;16:1939-49.
[23] Alam ME, Slaney TR, Wu L, Das TK, Kar S, Barnett GV*, et al.* Unique impacts of methionine oxidation, tryptophan oxidation, and asparagine deamidation on antibody stability and aggregation. Journal of Pharmaceutical Sciences. 2020;109:656-69.
[24] Vlasak J, Ionescu R. In: MAbs. Taylor & Francis; 2011. p. 253-63.
[25] Moritz B, Stracke JO. Assessment of disulfide and hinge modifications in monoclonal antibodies. Electrophoresis. 2017;38:769-85.
[26] Fan X, Brezski RJ, Fa M, Deng H, Oberholtzer A, Gonzalez A*, et al.* A single proteolytic cleavage within the lower hinge of trastuzumab reduces immune effector function and in vivo efficacy. Breast Cancer Research. 2012;14:1-13.
[27] Kulkarni JA, Witzigmann D, Thomson SB, Chen S, Leavitt BR, Cullis PR*, et al.* The current landscape of nucleic acid therapeutics. Nature nanotechnology. 2021;16:630-43.
[28] Rosa SS, Prazeres DM, Azevedo AM, Marques MP. mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine. 2021;39:2190-200.
[29] Voss C. Downstream processing of plasmid DNA for gene therapy and genetic vaccination. Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology. 2008;31:858-63.
[30] Moreira PI, Nunomura A, Nakamura M, Takeda A, Shenk JC, Aliev G*, et al.* Nucleic acid oxidation in Alzheimer disease. Free Radical Biology and Medicine. 2008;44:1493-505.
[31] Richardson B. Impact of aging on DNA methylation. Ageing research reviews. 2003;2:245-61.
[32] Barbieri I, Kouzarides T. Role of RNA modifications in cancer. Nature Reviews Cancer. 2020;20:303-22.

[33] Swenberg JA, Lu K, Moeller BC, Gao L, Upton PB, Nakamura J*, et al.* Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. Toxicological sciences. 2011;120:S130-S45.
[34] Pogocki D, Schöneich C. Chemical stability of nucleic acid–derived drugs. Journal of pharmaceutical sciences. 2000;89:443-56.
[35] Marmiy N, Esipov D. Biological role of 8-oxo-2'-deoxyguanosine. Moscow University biological sciences bulletin. 2015;70:168-72.
[36] Singal R, Ginder GD. DNA methylation. Blood, The Journal of the American Society of Hematology. 1999;93:4059-70.
[37] Krieg AM, Yi A-K, Matson S, Waldschmidt TJ, Bishop GA, Teasdale R*, et al.* CpG motifs in bacterial DNA trigger direct B-cell activation. Nature. 1995;374:546-9.
[38] Tóth R, Mészáros I, Hüser D, Forró B, Marton S, Olasz F*, et al.* Methylation status of the adeno-associated virus type 2 (AAV2). Viruses. 2019;11:38.
[39] Faust SM, Bell P, Cutler BJ, Ashley SN, Zhu Y, Rabinowitz JE*, et al.* CpG-depleted adeno-associated virus vectors evade immune detection. The Journal of clinical investigation. 2013;123:2994-3001.
[40] Frederico LA, Kunkel TA, Shaw BR. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry. 1990;29:2532-7.
[41] An R, Jia Y, Wan B, Zhang Y, Dong P, Li J*, et al.* Non-enzymatic depurination of nucleic acids: factors and mechanisms. PloS one. 2014;9:e115950.
[42] Calabretta A, Küpfer PA, Leumann CJ. The effect of RNA base lesions on mRNA translation. Nucleic acids research. 2015;43:4713-20.
[43] Sadeghipour F, Bugmann A, Herrera V, Bonnabry P. The reliability of operators when visually inspecting parenteral drugs. Pract Res Innov. 2007;13:41-5.
[44] Houde DJ, Berkowitz SA. Biophysical characterization and its role in the biopharmaceutical industry. In: Biophysical Characterization of Proteins in Developing Biopharmaceuticals. Elsevier; 2015.
[45] Skoog DA, Holler FJ, Crouch SR. Principles of instrumental analysis. Cengage learning; 2017.
[46] Cox R. Conformation of nucleic acids and the analysis of the hypochromic effect. Biochemical Journal. 1970;120:539.
[47] Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. Biotechniques. 1997;22:474-81.
[48] Weber A, Anderle H. Variable Pathlength Fiber-Optic Spectrophotometry for Protein Determination in Immunoglobulin Concentrates. BioPharm International. 2015;28:42–50-42–50.
[49] Compton SJ, Jones CG. Mechanism of dye response and interference in the Bradford protein assay. Analytical biochemistry. 1985;151:369-74.
[50] Brader ML. UV-absorbance, fluorescence and FT-IR spectroscopy in biopharmaceutical development. In: Biophysical Characterization of Proteins in Developing Biopharmaceuticals. Elsevier; 2020. p. 97-121.
[51] Fernández-Romero J, Aguilar-Caballos M. Fluorescence: Clinical and Drug Applications☆. 2018.
[52] Figeys D, Arriaga E, Renborg A, Dovichi NJ. Use of the fluorescent intercalating dyes POPO-3, YOYO-3 and YOYO-1 for ultrasensitive detection of double-stranded DNA separated by capillary electrophoresis with hydroxypropylmethyl cellulose and non-cross-linked polyacrylamide. Journal of Chromatography A. 1994;669:205-16.
[53] Michels DA, Brady LJ, Guo A, Balland A. Fluorescent derivatization method of proteins for characterization by capillary electrophoresis-sodium dodecyl sulfate with laser-induced fluorescence detection. Analytical chemistry. 2007;79:5963-71.
[54] Noble JE, Bailey MJ. Quantitation of protein. Methods in enzymology. 2009;463:73-95.

[55] Skoog DA, West DM, Holler FJ, Crouch SR. Fundamentals of analytical chemistry. Cengage learning; 2013.
[56] Sousa F, Sarmento B, Neves-Petersen MT. Biophysical study of bevacizumab structure and bioactivity under thermal and pH-stresses. European Journal of Pharmaceutical Sciences. 2017;105:127-36.
[57] Boulet-Audet M, Kazarian SG, Byrne B. In-column ATR-FTIR spectroscopy to monitor affinity chromatography purification of monoclonal antibodies. Scientific Reports. 2016;6:1-13.
[58] Salerno TMG, Coppolino C, Donato P, Mondello L. The online coupling of liquid chromatography to Fourier transform infrared spectroscopy using a solute-deposition interface: A proof of concept. Analytical and Bioanalytical Chemistry. 2022;414:703-12.
[59] Falke S, Betzel C. Dynamic Light Scattering (DLS) Principles, Perspectives, Applications to Biological Samples. Radiation in Bioanalysis: Spectroscopic Techniques and Theoretical Methods. 2019:173-93.
[60] Berne BJ, Pecora R. Dynamic light scattering: with applications to chemistry, biology, and physics. Courier Corporation; 2000.
[61] Xu R. Light scattering: A review of particle characterization applications. Particuology. 2015;18:11-21.
[62] Minton AP. Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. Analytical biochemistry. 2016;501:4.
[63] Stetefeld J, McKenna SA, Patel TR. Dynamic light scattering: a practical guide and applications in biomedical sciences. Biophysical reviews. 2016;8:409-27.
[64] Laue TM. Analytical ultracentrifugation. Current Protocols in Protein Science. 1996;4:7.5. 1-7.5. 9.
[65] Edwards GB, Muthurajan UM, Bowerman S, Luger K. Analytical ultracentrifugation (AUC): an overview of the application of fluorescence and absorbance AUC to the study of biological macromolecules. Current protocols in molecular biology. 2020;133:e131.
[66] Maeshima K, Rogge R, Tamura S, Joti Y, Hikima T, Szerlong H*, et al.* Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. The EMBO journal. 2016;35:1115-32.
[67] Burnham B, Nass S, Kong E, Mattingly M, Woodcock D, Song A*, et al.* Analytical ultracentrifugation as an approach to characterize recombinant adeno-associated viral vectors. Human Gene Therapy Methods. 2015;26:228-42.
[68] Doessegger L, Mahler H-C, Szczesny P, Rockstroh H, Kallmeyer G, Langenkamp A*, et al.* The potential clinical relevance of visible particles in parenteral drugs. Journal of pharmaceutical sciences. 2012;101:2635-44.
[69] Pham NB, Meng WS. Protein aggregation and immunogenicity of biotherapeutics. International journal of pharmaceutics. 2020;585:119523.
[70] Khodabandehloo A, Chen DDY. Particle sizing methods for the detection of protein aggregates in biopharmaceuticals. Bioanalysis. 2017;9:313-26.
[71] Das TK. Protein particulate detection issues in biotherapeutics development—current status. Aaps Pharmscitech. 2012;13:732-46.
[72] Garidel P, Herre A, Kliche W. Microscopic methods for particle characterization in protein pharmaceuticals. Analysis of aggregates and particles in protein pharmaceuticals. 2012:269-302.
[73] Ripple DC, Dimitrova MN. Protein particles: What we know and what we do not know. Journal of pharmaceutical sciences. 2012;101:3568-79.
[74] Kiyoshi M, Shibata H, Harazono A, Torisu T, Maruno T, Akimaru M*, et al.* Collaborative study for analysis of subvisible particles using flow imaging and light obscuration: experiences in Japanese biopharmaceutical consortium. Journal of Pharmaceutical Sciences. 2019;108:832-41.

[75] Franken LE, Grünewald K, Boekema EJ, Stuart MC. A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. Small. 2020;16:1906198.
[76] Brenner S, Horne R. A negative staining method for high resolution electron microscopy of viruses. Biochimica et biophysica acta. 1959;34:103-10.
[77] Sung JJ, Pardeshi NN, Mulder AM, Mulligan SK, Quispe J, On K*, et al.* Transmission electron microscopy as an orthogonal method to characterize protein aggregates. Journal of pharmaceutical sciences. 2015;104:750-9.
[78] Roingeard P, Raynal PI, Eymieux S, Blanchard E. Virus detection by transmission electron microscopy: Still useful for diagnosis and a plus for biosafety. Reviews in medical virology. 2019;29:e2019.
[79] Dastanpour R, Boone JM, Rogak SN. Automated primary particle sizing of nanoparticle aggregates by TEM image analysis. Powder Technology. 2016;295:218-24.
[80] McDonald K. High-pressure freezing for preservation of high resolution fine structure and antigenicity for immunolabeling. Electron microscopy methods and protocols. 1999:77-97.
[81] Gimpel AL, Katsikis G, Sha S, Maloney AJ, Hong MS, Nguyen TN*, et al.* Analytical methods for process and product characterization of recombinant adeno-associated virus-based gene therapies. Molecular Therapy-Methods & Clinical Development. 2021;20:740-54.
[82] Sanghani A, Kafetzis KN, Sato Y, Elboraie S, Fajardo-Sanchez J, Harashima H*, et al.* Novel PEGylated lipid nanoparticles have a high encapsulation efficiency and effectively deliver MRTF-B siRNA in conjunctival fibroblasts. Pharmaceutics. 2021;13:382.
[83] Brader ML, Williams SJ, Banks JM, Hui WH, Zhou ZH, Jin L. Encapsulation state of messenger RNA inside lipid nanoparticles. Biophysical journal. 2021;120:2766-70.
[84] Fu X, Chen W-C, Argento C, Clarner P, Bhatt V, Dickerson R*, et al.* Analytical strategies for quantification of adeno-associated virus empty capsids to support process development. Human Gene Therapy Methods. 2019;30:144-52.
[85] Gey MH. Instrumentelle Analytik und Bioanalytik: Biosubstanzen, Trennmethoden, Strukturanalytik, Applikationen. Springer-Verlag; 2015.
[86] Gross JH. Mass spectrometry: a textbook. Springer Science & Business Media; 2006.
[87] Chait BT. Mass spectrometry: bottom-up or top-down? Science. 2006;314:65-6.
[88] Resemann A, Wunderlich D, Rothbauer U, Warscheid B, Leonhardt H, Fuchser J*, et al.* Top-down de novo protein sequencing of a 13.6 kDa camelid single heavy chain antibody by matrix-assisted laser desorption ionization-time-of-flight/time-of-flight mass spectrometry. Analytical chemistry. 2010;82:3283-92.
[89] Valeja SG, Kaiser NK, Xian F, Hendrickson CL, Rouse JC, Marshall AG. Unit mass baseline resolution for an intact 148 kDa therapeutic monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. Analytical chemistry. 2011;83:8391-5.
[90] Qi C, Ding J, Yuan B, Feng Y. Analytical methods for locating modifications in nucleic acids. Chinese Chemical Letters. 2019;30:1618-26.
[91] Santos IC, Brodbelt JS. Recent developments in the characterization of nucleic acids by liquid chromatography, capillary electrophoresis, ion mobility, and mass spectrometry (2010–2020). Journal of separation science. 2021;44:340-72.
[92] Sandra K, Vandenheede I, Sandra P. Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. Journal of chromatography A. 2014;1335:81-103.
[93] Gahoual R, Busnel J-M, Beck A, François Y-N, Leize-Wagner E. Full antibody primary structure and microvariant characterization in a single injection using transient isotachophoresis and sheathless capillary electrophoresis–tandem mass spectrometry. Analytical Chemistry. 2014;86:9074-81.

[94] Fekete S, Guillarme D, Sandra P, Sandra K. Chromatographic, electrophoretic, and mass spectrometric methods for the analytical characterization of protein biopharmaceuticals. Analytical Chemistry. 2016;88:480-507.
[95] Schürch S. Characterization of nucleic acids by tandem mass spectrometry-The second decade (2004–2013): From DNA to RNA and modified sequences. Mass spectrometry reviews. 2016;35:483-523.
[96] Tretyakova N, Villalta PW, Kotapati S. Mass spectrometry of structurally modified DNA. Chemical reviews. 2013;113:2395-436.
[97] Harris RJ, Kabakoff B, Macchi FD, Shen FJ, Kwong M, Andya JD*, et al.* Identification of multiple sources of charge heterogeneity in a recombinant antibody. Journal of chromatography b: biomedical sciences and applications. 2001;752:233-45.
[98] Beckman JS, Voinov VG, Hare M, Sturgeon D, Vasil'ev Y, Oppenheimer D*, et al.* Improved protein and PTM characterization with a practical electron-based fragmentation on Q-TOF instruments. Journal of the American Society for Mass Spectrometry. 2021;32:2081-91.
[99] White JR, Abodeely M, Ahmed S, Debauve G, Johnson E, Meyer DM*, et al.* Best practices in bioassay development to support registration of biopharmaceuticals. BioTechniques. 2019;67:126-37.
[100] Layshock JR. Analytical Techniques for Biopharmaceutical Development—ELISA. Analytical Techniques for Biopharmaceutical Development. 2005:279.
[101] Besir H, Queitsch I, Hammer K, Odenwald C, Betts K, Kuhlendahl S. In: MOLECULAR THERAPY. CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA; 2019. p. 349-.
[102] Braun A, Alsenz J. Development and use of enzyme-linked immunosorbent assays (ELISA) for the detection of protein aggregates in interferon-alpha (IFN-α) formulations. Pharmaceutical research. 1997;14:1394-400.
[103] Anicetti VR, Fehskens EF, Reed BR, Chen AB, Moore P, Geier MD*, et al.* Immunoassay for the detection of E. coli proteins in recombinant DNA derived human growth hormone. Journal of Immunological Methods. 1986;91:213-24.
[104] Gombotz WR, Pankey SC, Phan D, Drager R, Donaldson K, Antonsen KP*, et al.* The stabilization of a human IgM monoclonal antibody with poly (vinylpyrrolidone). Pharmaceutical research. 1994;11:624-32.
[105] Golay J, Introna M. Mechanism of action of therapeutic monoclonal antibodies: promises and pitfalls of in vitro and in vivo assays. Archives of biochemistry and biophysics. 2012;526:146-53.
[106] Wang L, Yu C, Wang J. Development of reporter gene assays to determine the bioactivity of biopharmaceuticals. Biotechnology advances. 2020;39:107466.
[107] Luo Y, Matejic T, Ng C-K, Nunnally B, Porter T, Raso S*, et al.* Characterization and analysis of biopharmaceutical proteins. Separation science and technology. 2011;10:283-359.
[108] Garibyan L, Avashia N. Research techniques made simple: polymerase chain reaction (PCR). The Journal of investigative dermatology. 2013;133:e6.
[109] Teymouri M, Mollazadeh S, Mortazavi H, Ghale-Noie ZN, Keyvani V, Aghababaei F*, et al.* Recent advances and challenges of RT-PCR tests for the diagnosis of COVID-19. Pathology-Research and Practice. 2021;221:153443.
[110] Schochetman G, Ou C-Y, Jones WK. Polymerase chain reaction. The Journal of infectious diseases. 1988;158:1154-7.
[111] Joshi M, Deshpande J. Polymerase chain reaction: methods, principles and application. International Journal of Biomedical Research. 2010;2:81-97.
[112] Morrison TB, Weis JJ, Wittwer CT. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques. 1998;24:954-8, 60, 62.

[113] Taylor SC, Laperriere G, Germain H. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. Scientific reports. 2017;7:2409.
[114] Kojabad AA, Farzanehpour M, Galeh HEG, Dorostkar R, Jafarpour A, Bolandian M*, et al.* Droplet digital PCR of viral DNA/RNA, current progress, challenges, and future perspectives. Journal of Medical Virology. 2021;93:4182-97.
[115] Soltany-Rezaee-Rad M, Sepehrizadeh Z, Mottaghi-Dastjerdi N, Yazdi MT, Seyatesh N. Comparison of SYBR Green and TaqMan real-time PCR methods for quantitative detection of residual CHO host-cell DNA in biopharmaceuticals. Biologicals. 2015;43:130-5.
[116] Barczak W, Suchorska W, Rubiś B, Kulcenty K. Universal real-time PCR-based assay for lentiviral titration. Molecular biotechnology. 2015;57:195-200.
[117] Aurnhammer C, Haase M, Muether N, Hausl M, Rauschhuber C, Huber I*, et al.* Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. Human Gene Therapy, Part B: Methods. 2012;23:18-28.
[118] Baldwin J, Piplani S, Sakala IG, Honda-Okubo Y, Li L, Petrovsky N. Rapid development of analytical methods for evaluating pandemic vaccines: a COVID-19 perspective. Bioanalysis. 2021;13:1805-26.
[119] Gitlin JM, Institute NIoHNHGR. Calculating the economic impact of the Human Genome Project. 2012.
[120] Wadman M. Economic return from human genome project grows. Nature. 2013;10.
[121] Gibbs RA. The human genome project changed everything. Nature Reviews Genetics. 2020;21:575-6.
[122] Slatko BE, Gardner AF, Ausubel FM. Overview of next-generation sequencing technologies. Current protocols in molecular biology. 2018;122:e59.
[123] Bronner IF, Quail MA, Turner DJ, Swerdlow H. Improved protocols for illumina sequencing. Current protocols in human genetics. 2013;79:18.2. 1-.2. 42.
[124] Buermans H, Den Dunnen J. Next generation sequencing technology: advances and applications. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease. 2014;1842:1932-41.
[125] Woollard PM, Mehta NA, Vamathevan JJ, Van Horn S, Bonde BK, Dow DJ. The application of next-generation sequencing technologies to drug discovery and development. Drug discovery today. 2011;16:512-9.
[126] Morganti S, Tarantino P, Ferraro E, D'Amico P, Duso BA, Curigliano G. Next Generation Sequencing (NGS): a revolutionary technology in pharmacogenomics and personalized medicine in cancer. Translational Research and Onco-Omics Applications in the Era of Cancer Personal Genomics. 2019:9-30.
[127] Richards B, Cao S, Plavsic M, Pomponio R, Davies C, Mattaliano R*, et al.* Detection of adventitious agents using next-generation sequencing. PDA Journal of Pharmaceutical Science and Technology. 2014;68:651-60.
[128] MacDonald ML, Polson SW, Lee KH. k-mer-based metagenomics tools provide a fast and sensitive approach for the detection of viral contaminants in biopharmaceutical and vaccine manufacturing applications using next-generation sequencing. Msphere. 2021;6:e01336-20.
[129] Guerin K, Rego M, Bourges D, Ersing I, Haery L, Harten DeMaio K*, et al.* A novel next-generation sequencing and analysis platform to assess the identity of recombinant adeno-associated viral preparations from viral DNA extracts. Human gene therapy. 2020;31:664-78.
[130] Morris C, Lee YS, Yoon S. Adventitious agent detection methods in bio-pharmaceutical applications with a focus on viruses, bacteria, and mycoplasma. Current Opinion in Biotechnology. 2021;71:105-14.
[131] Mant CT, Hodges RS. High performance liquid chromatography of peptides and proteins. Separation, Analysis and Conformation. 1991.

[132] Andrews P. Estimation of the molecular weights of proteins by Sephadex gel-filtration. Biochemical Journal. 1964;91:222.
[133] Kastner M. Protein liquid chromatography. Elsevier; 1999.
[134] Krayukhina E, Uchiyama S, Nojima K, Okada Y, Hamaguchi I, Fukui K. Aggregation analysis of pharmaceutical human immunoglobulin preparations using size-exclusion chromatography and analytical ultracentrifugation sedimentation velocity. Journal of bioscience and bioengineering. 2013;115:104-10.
[135] Mant CT, Hodges RS. High performance liquid chromatography of peptides and proteins. Separation, Analysis and Conformation. 1991:2-4.
[136] Yan B, Valliere-Douglass J, Brady L, Steen S, Han M, Pace D*, et al.* Analysis of post-translational modifications in recombinant monoclonal antibody IgG1 by reversed-phase liquid chromatography/mass spectrometry. Journal of Chromatography A. 2007;1164:153-61.
[137] Werner A. Reversed-phase and ion-pair separations of nucleotides, nucleosides and nucleobases: analysis of biological samples in health and disease. Journal of Chromatography B: Biomedical Sciences and Applications. 1993;618:3-14.
[138] Huber CG, Oefner PJ, Bonn GK. Rapid and accurate sizing of DNA fragments by ion-pair chromatography on alkylated nonporous poly (styrene-divinylbenzene) particles. Analytical Chemistry. 1995;67:578-85.
[139] Wysoczynski CL, Roemer SC, Dostal V, Barkley RM, Churchill ME, Malarkey CS. Reversed-phase ion-pair liquid chromatography method for purification of duplex DNA with single base pair resolution. Nucleic acids research. 2013;41:e194-e.
[140] Yang Y, Wang G, Song T, Lebrilla CB, Heck AJ. In: MAbs. Taylor & Francis; 2017. p. 638-45.
[141] Chahar DS, Ravindran S, Pisal S. Monoclonal antibody purification and its progression to commercial scale. Biologicals. 2020;63:1-13.
[142] McCalley D. Hydrophilic-Interaction Chromatography: An Update. LCGC North America. 2020;38.
[143] Ikegami T. Hydrophilic interaction chromatography for the analysis of biopharmaceutical drugs and therapeutic peptides: A review based on the separation characteristics of the hydrophilic interaction chromatography phases. Journal of separation science. 2019;42:130-213.
[144] Lobue PA, Jora M, Addepalli B, Limbach PA. Oligonucleotide analysis by hydrophilic interaction liquid chromatography-mass spectrometry in the absence of ion-pair reagents. Journal of Chromatography A. 2019;1595:39-48.
[145] Cui T, Fakhfakh K, Turney H, Güler-Gane G, Toloczko A, Hulley M*, et al.* Comprehensive studies on building a scalable downstream process for mRNAs to enable mRNA therapeutics. Biotechnology Progress. 2022:e3301.
[146] Singh N, Heldt CL. Challenges in downstream purification of gene therapy viral vectors. Current Opinion in Chemical Engineering. 2022;35:100780.
[147] Rodriguez-Diaz R, Wehr T, Tuck S. Analytical techniques for biopharmaceutical development. Taylor & Francis US; 2005.
[148] Landers JP. Handbook of capillary and microchip electrophoresis and associated microtechniques. CRC press; 2007.
[149] Henley WH, Jorgenson JW. Ultra-high voltage capillary electrophoresis> 300 kV: recent advances in instrumentation and analyte detection. Journal of Chromatography A. 2012;1261:171-8.
[150] Bockris JM, Devanathan M, Müller K. On the structure of charged interfaces. In: Electrochemistry. Elsevier; 1965. p. 832-63.
[151] Stutz H. Protein attachment onto silica surfaces–a survey of molecular fundamentals, resulting effects and novel preventive strategies in CE. Electrophoresis. 2009;30:2032-61.
[152] Murashov VV, Leszczynski J. Adsorption of the phosphate groups on silica hydroxyls: an ab initio study. The Journal of Physical Chemistry A. 1999;103:1228-38.

[153] Shi B, Shin YK, Hassanali AA, Singer SJ. DNA binding to the silica surface. The Journal of Physical Chemistry B. 2015;119:11030-40.
[154] Beckers JL, Boček P. The preparation of background electrolytes in capillary zone electrophoresis: Golden rules and pitfalls. Electrophoresis. 2003;24:518-35.
[155] Zhang X, McGown LB. Sequence-based separation of single-stranded DNA at high salt concentrations in capillary zone electrophoresis. Electrophoresis. 2016;37:2017-24.
[156] Moritz B, Schnaible V, Kiessig S, Heyne A, Wild M, Finkler C*, et al.* Evaluation of capillary zone electrophoresis for charge heterogeneity testing of monoclonal antibodies. Journal of Chromatography B. 2015;983:101-10.
[157] He Y, Isele C, Hou W, Ruesch M. Rapid analysis of charge variants of monoclonal antibodies with capillary zone electrophoresis in dynamically coated fused-silica capillary. Journal of separation science. 2011;34:548-55.
[158] Pajaziti B, Petkovska R, Andrási M, Nebija D. Application of the capillary zone electrophoresis (CZE) and capillary gel electrophoresis (CGE) for the separation of human insulin, insulin lispro and their degradation products. Die Pharmazie-An International Journal of Pharmaceutical Sciences. 2020;75:167-71.
[159] Stolz A, Jooß K, Höcker O, Römer J, Schlecht J, Neusüß C. Recent advances in capillary electrophoresis-mass spectrometry: Instrumentation, methodology and applications. Electrophoresis. 2019;40:79-112.
[160] Han M, Rock BM, Pearson JT, Rock DA. Intact mass analysis of monoclonal antibodies by capillary electrophoresis—Mass spectrometry. Journal of Chromatography B. 2016;1011:24-32.
[161] Ebersole RC, McCormick RM. Separation and isolation of viable bacteria by capillary zone electrophoresis. Bio/technology. 1993;11:1278-82.
[162] Neubert RH, Ruttinger H-H. Affinity capillary electrophoresis in pharmaceutics and biopharmaceutics. CRC Press; 2003.
[163] Villareal V, Kaddis J, Azad M, Zurita C, Silva I, Hernandez L*, et al.* Partial-filling affinity capillary electrophoresis. Analytical and bioanalytical chemistry. 2003;376:822-31.
[164] Wang Y, Adeoye DI, Ogunkunle EO, Wei I-A, Filla RT, Roper MG. Affinity capillary electrophoresis: A critical review of the literature from 2018 to 2020. Analytical Chemistry. 2020;93:295-310.
[165] Hutanu A, Kiessig S, Bathke A, Ketterer R, Riner S, Olaf Stracke J*, et al.* Application of affinity capillary electrophoresis for charge heterogeneity profiling of biopharmaceuticals. Electrophoresis. 2019;40:3014-22.
[166] Gstöttner C, Hook M, Christopeit T, Knaupp A, Schlothauer T, Reusch D*, et al.* Affinity Capillary Electrophoresis–Mass Spectrometry as a Tool to Unravel Proteoform-Specific Antibody–Receptor Interactions. Analytical Chemistry. 2021;93:15133-41.
[167] Lin J, Lazar AC. Determination of charge heterogeneity and level of unconjugated antibody by imaged cIEF. Antibody-Drug Conjugates. 2013:295-302.
[168] Salas-Solano O, Kennel B, Park SS, Roby K, Sosic Z, Boumajny B*, et al.* Robustness of i CIEF methodology for the analysis of monoclonal antibodies: an interlaboratory study. Journal of separation science. 2012;35:3124-9.
[169] Mack S, Darling S. Mastering charge heterogeneity analysis of therapeutic proteins: Understanding crucial aspects of icIEF method development and implementation. Genetic Engineering & Biotechnology News. 2014;34:38-9.
[170] Dai J, Lamp J, Xia Q, Zhang Y. Capillary isoelectric focusing-mass spectrometry method for the separation and online characterization of intact monoclonal antibody charge variants. Analytical chemistry. 2018;90:2246-54.
[171] Karger BL, Cohen AS, Guttman A. High-performance capillary electrophoresis in the biological sciences. Journal of Chromatography. 1989;492:585-614.

[172] Sänger–van de Griend CE. CE-SDS method development, validation, and best practice—An overview. Electrophoresis. 2019;40:2361-74.
[173] Heller C. Capillary electrophoresis of proteins and nucleic acids in gels and entangled polymer solutions. Journal of Chromatography A. 1995;698:19-31.
[174] Scheller C, Krebs F, Wiesner R, Wätzig H, Oltmann-Norden I. A comparative study of CE-SDS, SDS-PAGE, and Simple Western—Precision, repeatability, and apparent molecular mass shifts by glycosylation. Electrophoresis. 2021;42:1521-31.
[175] Sánchez-Hernández L, Montealegre C, Kiessig S, Moritz B, Neusüß C. In-capillary approach to eliminate SDS interferences in antibody analysis by capillary electrophoresis coupled to mass spectrometry. Electrophoresis. 2017;38:1044-52.
[176] García-Cañas V, González R, Cifuentes A. Ultrasensitive detection of genetically modified maize DNA by capillary gel electrophoresis with laser-induced fluorescence using different fluorescent intercalating dyes. Journal of Agricultural and Food Chemistry. 2002;50:4497-502.
[177] Berns KI, Giraud C. Adeno-associated virus (AAV) vectors in gene therapy. Springer Science & Business Media; 2012.
[178] Shakeel S, Karim S, Ali A. Peptide nucleic acid (PNA)—a review. Journal of Chemical Technology & Biotechnology: International Research in Process, Environmental & Clean Technology. 2006;81:892-9.
[179] Vogt W. Oxidation of methionyl residues in proteins: tools, targets, and reversal. Free Radical Biology and Medicine. 1995;18:93-105.
[180] Stracke J, Emrich T, Rueger P, Schlothauer T, Kling L, Knaupp A*, et al.* In: MAbs. Taylor & Francis; 2014. p. 1229-42.
[181] Zhang Z, Pan H, Chen X. Mass spectrometry for structural characterization of therapeutic antibodies. Mass spectrometry reviews. 2009;28:147-76.
[182] Wang D, Tai PW, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. Nature reviews Drug discovery. 2019;18:358-78.
[183] Fekete S, Aebischer MK, Imiołek M, Graf T, Ruppert R, Lauber M*, et al.* Chromatographic strategies for the analytical characterization of adeno-associated virus vector-based gene therapy products. TrAC Trends in Analytical Chemistry. 2023:117088.
[184] Wright JF. Quality control testing, characterization and critical quality attributes of adeno-associated virus vectors used for human gene therapy. Biotechnology Journal. 2021;16:2000022.
[185] Perry-O'Keefe H, Yao XW, Coull JM, Fuchs M, Egholm M. Peptide nucleic acid pre-gel hybridization: an alternative to southern hybridization. Proc Natl Acad Sci U S A. 1996;93:14670-5.
[186] Gupta A, Mishra A, Puri N. Peptide nucleic acids: Advanced tools for biomedical applications. Journal of biotechnology. 2017;259:148-59.
[187] Basile A, Giuliani A, Pirri G, Chiari M. Use of peptide nucleic acid probes for detecting DNA single-base mutations by capillary electrophoresis. Electrophoresis. 2002;23:926-9.

5. Abbreviations

AEX	anion-exchange
AF4	Asymmetric flow field flow fractionation
ALC	Affinity LC
API	active pharmaceutical ingredient
ASO	antisense oligonucleotides
AUC	analytical ultracentrifugation
CD	Circular dichroism
CE	capillary electrophoresis
CEX	cation-exchange
C&GT	cell and gene therapy
CGE	capillary gel electrophoresis
CID	collision-induced dissociation
CQA	critical quality attributes
ddPCR	digital droplet PCR
DLS	dynamic light scattering
ds	double stranded
DSC	Differential scanning calorimetry
EGFP	enhanced green fluorescent protein
ELISA	enzyme-linked immunosorbent assay
ESI	electrospray ionization
FTIR	fourier-transform-IR
HIC	hydrophobic interaction chromatography
HILIC	hydrophilic interaction liquid chromatography
HMW	high molecular weight species
HPLC	high pressure/performance liquid chromatography
HTS	high throughput sequencing
IEC	ion exchange chromatography
IP-RP-HPLC	ion pair-RP-HPLC
IR	infrared
LC	liquid chromatography
LNP	lipid nano particle
LOQ	limit of quantification
mAb	monoclonal antibody
MALS	multi angle light scattering
MEKC	Micellar electrokinetic chromatography
MS	mass spectrometry
M_w	molecular weight
NA	nucleic acid
NGS	next generation sequencing
NP	normal phase
nt	nucleotides
PCR	polymerase chain reaction
pDNA	plasmid DNA
pI	isoelectric point
PNA	peptide nucleic acid
PTM	post-translational modifications
QC	quality control
qPCR	quantitative PCR
rAAV	recombinant adeno associated virus
RP	revered phase
RT	reverse transcriptase
SEC	size exclusion chromatography
siRNA	small interfering RNA
SPR	Surface plasmon resonance
ss	single stranded
UV	Ultraviolet (UV),
VIS	visible
TEM	transmission electron microscopy
TFA	trifluoroacetic acid

6. Acknowledgements

I would like to express my deepest gratitude to my supervisor, PD Dr. Maria Anna Schwarz, for her guidance, encouragement and support throughout my doctoral studies. She has been a mentor, a friend and an inspiration to me. I am indebted to her for her invaluable feedback, constructive criticism and insightful suggestions that helped me improve the quality of my research and writing. I will greatly miss our bi-weekly afternoon meeting with the delicious tea.

I am also grateful to my co-supervisors, Dr. Bernd Moritz and Dr. Steffen Kiessig, for enabling this research at F. Hoffmann-La-Roche Ltd., Basel. They have been very supportive and helpful in all aspects of my research. Their profound expertise and detailed advice have been very beneficial and are much appreciated.

I would like to thank the members of my thesis committee, Prof. Dr. Peter Hauser from the University of Basel and Prof. Dr. Hermann Wätzig from the Technical University Braunschweig for their time and interest in reviewing my work.

I would like to acknowledge the role of Dr. Adelheid Rohde, Dr. Jan Olaf Stracke and Dr. Markus Wild. They have provided me with the necessary resources, facilities and opportunities to conduct my research at F. Hoffmann-La-Roche Ltd., Basel. I appreciate their trust and confidence in me.

My colleagues and friends at PTDEA and Roche and especially the CE team have been a great source of camaraderie, collaboration and assistance. I have learned a lot from them and enjoyed working with them. Dr. Manuel Gregoritza greatly supported the PNA publication with his constructive feedback. I am especially thankful to Andrea Bathke, Rolf Ketterer and Sonja Riner who taught me how to handle delicate CE devices and were always there when troubleshooting was necessary.

I would like to thank the three interns I had the pleasure of supervising and who helped me out in the lab: Aurélie Noël, Mike Dang Thai, and Chiara Signori. They have been dedicated, eager, and dependable and made substantial contributions to the collection and analysis of data for my studies. It has been enjoyable to collaborate with them and I am thankful for their invaluable feedback that helped me to develop my communication and management skills.

I would like to dedicate a section to my colleagues from Solvias AG especially Dr. Cristina Montealegre, Dr. Marius Koch, Dr. Balazs Bobaly, Dr. Oliver Höcker and Dr. Claudio Schmidli. Through sharing data, expertise and insights with me we were able to publish two papers. They have been great partners and collaborators.

A special thanks goes to my friend Dr. Christoph Gstöttner from Leiden University. He sparked the idea of a joint effort for the IP-RP-LC publication and has been a wonderful collaborator.

I also want to thank the numerous people, who will go unnamed in this appreciation, but assisted me in many ways throughout my research. They have enriched this work with their perspectives, knowledge and skills. I thank everyone for their efforts and apologize for any omission or oversight.

My profound gratitude goes to Alena Freudenmann, who has been my rock and inspiration throughout this journey. She has often had the best project ideas for my research and has always encouraged me to pursue the work. Thank you for being patient, understanding and supportive of me and my work.

Last but not least, I would like to dedicate this thesis to my family: my parents Natalia and Vladimir and my brother Alexander. They have sacrificed a lot for me and supported me unconditionally. I could not have achieved this without them.

7. Curriculum Vitae: Andrei Hutanu

Education

2012 – 2015	**Bachelor of Science in Biochemistry** at the Technical University of Munich (TUM) Bachelor's Thesis: "Characterization Of Halophile Microalgae And Their Process Optimization in A Downscaled Setup"
2015 – 2018	**Master of Science in Biochemistry** at the Technical University of Munich (TUM) Master's Thesis: "Application of Affinity Capillary Electrophoresis for Charge Heterogeneity Profiling of Biopharmaceuticals"
Since 2018	**Doctor of Philosophy (Ph.D.)** at the University of Basel

Work Experience

Summer 2015	**Internship** at Clariant AG, Planegg, Germany Assay development for cellulase-activity measurements.
2016 – 2017	**Internship** at F. Hoffmann - La Roche AG, Basel, Switzerland Method development for slab gel electrophoresis and offline MS coupling
2017 – 2018	**Master Thesis Student** at F. Hoffmann - La Roche AG, Basel, Switzerland Implementation of ACE for the analysis of Antibody Co-formulations.
2018-2022	**Ph.D. Researcher** at F. Hoffmann - La Roche AG, Basel, Switzerland
Since 2023	**Fellow Analytical Development** at Ten23 health AG, Basel, Switzerland

List of Publications

HUTANU A, KIESSIG S, BATHKE A, KETTERER R, RINER S, OLAF STRACKE J, WILD M, MORITZ B
Application of affinity capillary electrophoresis for charge heterogeneity profiling of biopharmaceuticals.
Electrophoresis. 2019 Nov;40(22):3014-22.

HUTANU A, HAUSER PC, MORITZ B, KIESSIG S, NOËL A, STRACKE JO, WILD M, SCHWARZ MA
Methionine oxidation of proteins analyzed by affinity capillary electrophoresis in presence of silver (I) and gold (III) ions.
Electrophoresis. 2021 Jun;42(11):1209-16.

HUTANU A, BOELSTERLI D, SCHMIDLI C, MONTEALEGRE C, DANG THAI MH, BOBALY B, KOCH M, SCHWARZ MA

Stronger together: Analytical techniques for recombinant adeno associated virus.

Electrophoresis. 2022 May;43(9-10):1107-17.

HUTANU A, SIGNORI C, MORITZ B, GREGORITZA M, ROHDE A, SCHWARZ MA

Using Peptide Nucleic Acid Hybridization Probes for Qualitative and Quantitative Analysis of Nucleic Acid Therapeutics by Capillary Electrophoresis.

Analytical Chemistry. 2023 Mar;95(11):4914-22.

GSTÖTTNER C*, HUTANU A*, BOON S, RADUCANU A, RICHTER K, HAINDL M, RUPPERT R, DOMÍNGUEZ-VEGA E

**shared first authors*

Reversed Phase-Liquid Chromatography for Recombinant AAV Genome Integrity Assessment.

Analytical Chemistry. 2023 May; 95(22):8478-86.

HUTANU A, LÓPEZ FERREIRO M, VAN HAASTEREN J, HÖCKER O, MONTEALEGRE C, MÄSER M, KERESZTFALVI A, MONTI J, SCHWARZ MA

Electrophoretic characterization of LNP/AAV encapsulated nucleic acids: Strengths and weaknesses.

Electrophoresis. 2023 Oct; 44(19-20):1595-606