American Journal of Biochemistry and Biotechnology

Review Article

Assessing Protein Purity

Shiva Siddappa[1], Liang-Yin Ke[4], Chu-Huang Chen[5] and Gopal Kedihithlu Marathe[2,3]

[1]Division of Biochemistry, School of Life Sciences, JSS Academy of Higher Education and Research, Mysuru 570015, Karnataka, India
[2]Department of Studies in Biochemistry, University of Mysore, Manasagangothri, Mysuru, 570006, Karnataka, India
[3] Department of Studies in Molecular Biology, University of Mysore, Manasagangothri, Mysuru, 570006, Karnataka, India
[4]Department of Medical Laboratory Science and Biotechnology, College of Health Sciences, KMU, Kaohsiung, Taiwan
[5]Vascular and Medicinal Research, Texas Heart Institute, Houston, Texas, 77030, USA

Article history
Received: September 3, 2024
Revised: December 21, 2024
Accepted: December 30, 2024

Corresponding Author:
Shiva Siddappa
Division of Biochemistry, School of Life Sciences, JSS Academy of Higher Education and Research, Mysuru 570015, Karnataka, India.
Email: shivasiddappa@jssuni.edu.in

Abstract: Determination of protein purity is crucial, as proteins are frequently used as therapeutic agents and reagents in many biochemical reactions. Additionally, purified protein is necessary for biophysical and structural studies. Assessing protein purity involves examining for impurities or demonstrating their absence. In our experience, a satisfactory answer is seldom obtained when biochemistry students are asked a common question on how to establish the purity of a given protein. This is not surprising, as no single technique firmly establishes protein purity. A combination of several techniques/approaches is needed for the assessment of protein purity.
SDS-PAGE, 2-dimensional electrophoresis, mass spectral methods, and N-terminal sequencing are the front-line techniques used in the assessment of protein purity. Ignoring protein purity can result in erroneous interpretations in experimental results. This review describes the diverse methods employed to confirm protein purity.

Keywords: Protein Purity, SDS-PAGE, 2-Dimensional Electrophoresis, Purity Assessment, N-terminal Sequencing

Introduction

Proteins serve as the fundamental building blocks and are vital for all the biochemical processes that occur within an organism. The proteome of a cell undergoes constant changes and systematic analysis is crucial for understanding their functions. Determining the purity of proteins can be achieved by various methods, including electrophoresis, N-terminal sequencing, chromatography, mass spectrometry, light scattering, or testing enzyme activity to a constant specific activity (Fig. 1) (Fowler *et al*., 1996; Rhodes and Laue, 2009).

More than a century ago, the famous bacteriologist Robert Koch (1843-1910) (Tan and Berman, 2008) advised that to study the pathogenesis of an infectious disease it is important to identify the causative agent and isolation of the causative microorganism in its pure form. Likewise, to study a reaction/function catalyzed by a particular protein, homogeneously purifying the protein under study is essential. Each chemical event inside a cell depends on the action of at least one specific protein (or enzyme). Understanding the nature and functions of a single protein is essential to study its role in physiological and pathological settings (De Souza, 2013; Westermarck *et al*., 2013). Only purified protein is required for determining amino acid sequence, establishing evolutionary relationships between proteins in diverse organisms, and correlating biochemical function to structural and biological aspects. Moreover, highly purified protein is essential for producing antibodies. Hence, purifying protein to homogeneity is vital in any biochemical analysis.

A popular fundamental mantra/commandment of enzymology is, “don’t waste clean thinking on dirty enzymes” (Admonition of Efraim Racker’s –1950s) (Kornberg, 2003). The saying simply means that a crude extract contains a mixture of many different enzymes, and performing a detailed study on enzyme function (conversion of substrate into product) requires a pure enzyme. It’s a waste of time if the enzyme is not purified to homogeneity (Kornberg, 2009). Even though SDS-PAGE and 2-dimensional electrophoresis are cutting-edge procedures, the data from these techniques alone cannot be used to assess the purity of the protein.

Therefore, additional techniques like HPLC, SEC, N-terminal sequencing, Microfluidic Diffusional Sizing (MDS), Dynamic Light Scattering (DLS), and mass spectrometry must be used to evaluate the purity and homogeneity of the purified protein. In conclusion, to date, no single method can be used to confirm that a protein under study is pure; rather, a combination of several techniques is required (Berrow *et al*., 2021; de Marco *et al*., 2021; Pirie, 1940; Raynal *et al*., 2014; Shedlovsky, 1943).

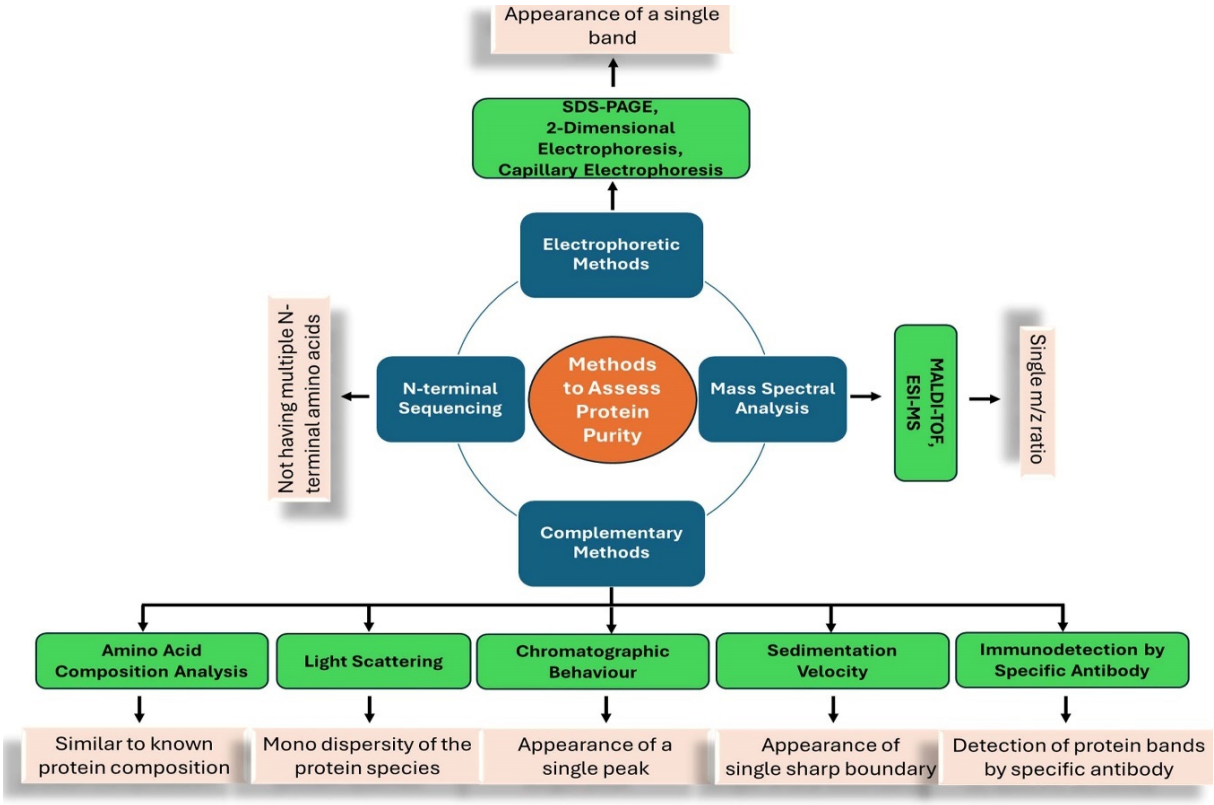

Fig. 1: Various analytical methods used to assess protein purity

In this review, we discuss the different analytical techniques that are generally used to establish protein purity (Fig. 1).

Determination of Protein Purity Using Electrophoretic Methods

Polyacrylamide Gel Electrophoresis (PAGE) is often the first method employed in assessing protein purity due to its low cost and ease of analysis. This approach separates protein mixtures based on charge, charge/mass, size, or shape. The popular electrophoretic techniques include sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), two-dimensional electrophoresis (2-DE), and Capillary Electrophoresis (CE).

Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE is the frontline method used for separation, detection, and determination of the purity of protein (Fig. 3). Till today, SDS-PAGE is widely used to demonstrate the purity of purified protein. Recently, Guerin *et al*., Salarifar *et al*., Rajabi *et al*., and Pechlivani, *et al*., analyzed the purity of recombinant proteins by SDS-PAGE (Guérin *et al*., 2024; Pechlivani *et al*., 2022; Rajabi *et al*., 2022; Salarifar *et al*., 2023). The purity of affinity-Purified Polyphenol Oxidase (PPO) from fruits and vegetables was assessed by SDS-PAGE (Öztürk and Küfrevioğlu, 2024). SDS-PAGE typically performed using Laemmli's protocol (Laemmli, 1970), separates the charged proteins according to their molecular mass. SDS-PAGE resolves impurities that have a different molecular mass. However, if the contaminant protein molecular mass is same as protein of interest, then using non-denaturing gel electrophoresis (native PAGE) or Isoelectric Focusing (IEF) would be better for separating the impurity.

A simple criterion is often used to establish protein purity: the appearance of a single band after electrophoresis. However, as stated above, this criterion cannot be used in all cases. If two or more similar molecular weight proteins are present along with the

protein of interest, they will run together as a single band on SDS-PAGE (Mohan, 1992). This problem can be resolved by adopting any one of the following methods:

1. Analyze the banding pattern further in 2-DE
2. Conduct the activity assay in the presence and absence of specific inhibitors (if both proteins are enzymes and enzyme activity for each can be inhibited by separate specific inhibitors)
3. Use subjects that have a mutation for one type of the enzyme/protein

Marathe *et al.* (2003) illustrated the above condition by studying Platelet Activating Factor Acetyl-Hydrolase (PAF-AH) and Paraxonase-1 (PON-1) (Marathe *et al*., 2003). PAF-AH is a Ca^{2+}-independent phospholipase A_2 that hydrolyses *Sn-2* residue of biologically active ether lipid PAF, whereas PON-1 is a Ca^{2+}-dependent esterase associated with high-density lipoprotein with uncertain function(s). Although it has been claimed that the PON-1 enzyme can hydrolyze PAF in a manner similar to PAF-AH (Rodrigo *et al*., 2001), Marathe *et al*., convincingly showed that trace amounts of co-purifying PAF-AH with PON-1 (both have a similar molecular weight) were responsible for the observed hydrolysis of PAF (Marathe *et al*., 2003) and that PON-1 itself was devoid of the ability to hydrolyze PAF (Watson *et al*., 1995). Just trace amounts of PAF-AH co-eluting with PON-1 misled the entire scientific community that PON-1 could hydrolyze PAF (Rodrigo *et al*., 2001). Moreover, these authors showed that subjects deficient in PAF-AH activity had PON-1 activity but were incapable of hydrolyzing PAF. They used two specific inhibitors such as EDTA (a Ca^{2+} chelator) and 4-(2-Aminoethyl) benzenesulfonyl fluoride (a specific serine protease inhibitor) that effectively inhibited PON-1 and PAF-AH, respectively. Finally, Marathe *et al* showed that recombinant PON1 enriched with esterase was devoid of PAF-AH activity. In conclusion, the authors were able to demonstrate that the hydrolysis of PAF is due to co-elution of trace amounts of PAF-AH along with PON-1.

Two-Dimensional Electrophoresis (2-DE)

2-DE is a superior analytical technique to SDS-PAGE and can help in separating protein of interest from closely related contaminants/impurities (Fig. 3). IEF and SDS-PAGE are used in 2-DE separation, and the modified O'Farrell (1975) method is typically used in analyzing proteins (O'Farrell, 1975). This technique allows investigators to resolve proteins of similar mass, if the net charge of the contaminant protein(s) differs from that of the protein of interest and can also be used to separate post-translationally modified proteins (eg, phosphorylated, glycosylated, lipidated, and sulphated proteins). In the first dimension of 2D-electrophoresis, various techniques can be employed, including conventional IEF, Immobilized pH Gradient (IPG) and non-equilibrium pH gel electrophoresis (NEPHGE) (Magdeldin *et al*., 2014). IPG strips are commonly used because they are mechanically stronger and avoid cationic accumulation. Separation of basic proteins using IPG strips results in severe loss of the protein sample with less reproducibility; therefore, the NEPHGE technique was developed and seems to have better resolution and spot reproducibility of basic proteins (Slibinskas *et al*., 2013). 2-DE has several advantages, such as its robustness, visual mapping (the ability to resolve more than 5000 proteins in a single gel) of proteins, and its compatibility with downstream analysis. However, the separation of hydrophobic, extremely acidic, and basic proteins is difficult in 2-DE. In addition, low reproducibility, low throughput, and high labor demands are some of the important drawbacks of 2DE (Görg *et al*., 1988; Rabilloud *et al*., 2010).

Zannis *et al*. identified four subunits of the same protein in 2-DE that appear as a single band in SDS-PAGE. The PI value of each subunit is different (6.63, 6.41, 6.29 and 6.20 for subunit 1-4, respectively), denoting the presence of more negative charges in subunits 2, 3 and 4 than in subunit 1 (Zannis *et al*., 1978). This example explains the scope of 2-DE in protein purity assessment.

2DE is frequently coupled with mass spectrometry for both qualitative and quantitative analysis. For instance, the quality of the recombinant protein, abatacept, was assessed by Nebija *et al* using both 2DE and MALDI-TOF. This combined approach enabled them to assess the critical attributes such as purity and identity of the product, charge heterogeneity, isoform pattern, and post-translational modifications (Nebija *et al*., 2011).

Capillary Electrophoresis (CE)

Capillary electrophoresis is well-suited for protein characterization because of its superior separation efficiency and requirement of small sample volumes. Due to the flexibility and feasibility of CE, different separation modes and detection methods can be adapted when using this technique. These methods include the following:

1. Capillary Isoelectric Focusing (CIEF) – molecules are separated based on their isoelectric point
2. SDS Capillary Gel Electrophoresis (SDS-CGE) protein separation is based on size or hydrodynamic radius

3. Capillary Zone Electrophoresis (CZE) protein separation purely depends on charge and hydrodynamic radius of the analyte
4. Capillary Electrophoresis-Mass Spectrometry (CE-MS) protein separation and molecular mass detection in a single analysis

With its better resolution, sensitivity, reproducibility, on-column detection, and automated operation, CE overcomes many of the drawbacks of conventional protein separation techniques such as SDS-PAGE and 2-DE. Significant limitation of capillary electrophoresis is the protein adsorption to fused silica capillary walls, which leads to inconsistent migration times and increased band broadening (Raynal *et al*., 2014; Whitmore and Gennaro, 2012; Zhao and Chen, 2014; Zhu *et al*., 2012).

One excellent example of the sensitivity of CE is the identification of two structural isoforms of IgG2. Non-reduced SDS-CGE was developed by Guo *et al*. for the separation of two structural isoforms of IgG2, and they described the disulfide heterogeneity in IgG2 antibodies (Guo *et al*., 2008).

Another important example involves the detection of low-mass peptides by CE-MS. Whitmore *et al*. analyzed the percent sequence coverage by both CE-MS and LC-MS for an in-house manufactured monoclonal antibody; they could detect 97% sequence coverage by LC-MS. The 3% not covered consisted of 11 small peptides that were easily detected by CE-MS with 100% sequence coverage (Whitmore *and* Gennaro, 2012). Hence amalgamation of CE with MS is a promising bio-analytical technique (Kumar *et al*., 2022). These examples illustrate the sensitivity of CE and its application in the analysis of protein purity.

As explained above, the most advanced techniques of gel electrophoresis and its flexibility in combining with other techniques, gel electrophoresis is the first and better method of choice for assessing protein purity. However, several potential problems with these techniques are worth noting:

1. If a contaminant co-elutes with the protein of interest but cannot penetrate the gel, then the entire stacking and resolving portions of the gel have to be stained to examine for the presence of the contaminant in the loading well or between the stacking and resolving gel. A band in the stacking gel indicates a high molecular weight or less soluble contaminant
2. Protein bands in the gel are mainly visualized by various staining methods. In most laboratories, silver staining and CBB are the routinely used staining protocols. Although silver staining is more sensitive than CBB, it is not compatible with further analysis of protein by mass spectroscopy. Although CBB is compatible with MS analysis of proteins, it's binding to fibrous proteins or glycoproteins is weak
3. Detection of contaminants in a sample depends on the concentration of each contaminant. Usually, the gel is loaded with a high amount of protein for the detection of contaminants. However, with higher sample concentrations and large sample volumes, distortion or band broadening may occur, which may result in similar electrophoretic mobility of impurities with the protein of interest
4. The recovery of protein in its native form becomes much difficult if the protein is treated with detergents such as SDS. The extent of renaturation depends largely on protein structure; a multimeric protein is less likely than a monomer to return to its native form
5. Artifacts, which occur during protein sample preparations and electrophoresis, may mislead the interpretation of the original banding pattern of the protein of interest

N-terminal Sequencing in Assessing Protein Purity

The protein sequencing critically relies on starting with a highly pure protein for successful sequencing. In 1950, Pehr Edman developed a chemical-based method popularly referred to as "Edman degradation" in which phenylisothiocyanate (PITC) derivatized amino acids are sequentially removed from the N-terminal end of the protein and identified in the order of their occurrence (Edman, 1950). However, N-terminal chemical blockage can occur in many naturally occurring proteins (50-80% of all eukaryotic proteins); in those cases, internal or C-terminal sequencing is useful. Several modern automated methods have been developed for N-terminal sequencing and can detect >30 residues in one run with less sample consumption (10-20 picomoles of purified protein). The development of more advanced techniques, such as mass spectrometry, substantially reduced the demand for traditional N-terminal sequencing (Speicher *et al*., 2009). Nevertheless, N-terminal sequencing is the method of choice for protein characterization. For example, Stern *et al*. purified an unknown cytokine from a human B-lymphoblastoid cell line. N-terminal analysis of this purified protein showed it was not related to any previously identified cytokine (Stern *et al*., 1990).

The free N-terminus of a protein is coupled with PITC in Edman degradation. However, proteins with N-terminal blockage may naturally occur (acetyl, formyl, or pyroglutamyl groups). During protein isolation or sample

preparation, some N-terminal modifications occur. For example, acrylamide monomer can react with the free N-termini of the protein. Urea is a reagent frequently used in protein purification procedures that can lead to protein modification. Although urea does not react directly with proteins, it decomposes readily to cyanate, which can react with free amino groups of proteins (Speicher *et al*., 2009). N-terminal blockage prevents the proteins from coupling with PITC. Certain methods can be used to deblock the N-terminal blockage of proteins (Chiari *et al*., 1992; Fowler *et al*., 1996; Speicher *et al*., 2009; Wellner *et al*., 1990). However, none of the deblocking methods is very effective, and the method used varies from protein to protein, depending on the nature of the N-terminal blockage.

The purity of the protein can be assessed by analyzing the presence of only one amino acid per hydrolysis (Fig.3). If the protein of interest is contaminated with other proteins in equal abundance, more than two amino acids are detected per hydrolysis (Fig. 2). Hence, N-terminal sequencing more precisely tells us whether the protein under study is pure or not. One potential problem in N-terminal sequencing is that contaminants often present in trace amounts cannot be detected. However, N-terminal sequencing of proteins has its purview over modern mass spectral analysis of proteins, even though sample deblocking is challenging.

Assessment of Protein Purity by Mass Spectral Methods

Until 1970, several analytical techniques such as electrophoresis, chromatography, and ultracentrifugation were the only available methods for estimating the molecular mass of proteins. Now, characterizing proteins beyond their molecular mass is essential in modern science; therefore, mass spectrometry is a better method of choice. Mass spectrometry detects the presence of any impurity(s) in the sample and provides the molecular mass with 0.01% accuracy, even with small amounts of a protein sample (picomole). In addition, it is possible to characterize the impurity by its mass and to identify the origin of the impurity, which may provide clues for avoiding such impurities (Tipton *et al*., 2011). The emergence of advanced ionization methods like Electrospray Ionization (ESI) and Matrix-Assisted Laser Desorption/Ionization (MALDI), when combined with Time-Of Flight (TOF) analyzers, significantly expanded the use of mass spectrometry in biological research. Specifically, MALDI-TOF enables the precise determination of molecular weight of proteins and the protein purity can be assessed by observing the presence of single m/z ratio (Fig. 3). For example, Guerin *et al* determined the mass of recombinant proteins CspZ and FhbA by Mass Spectrometry (MS) (Guerin *et al*., 2024). Pechlivani *et al* assessed the purity of purified fibrinogen by mass spectrometry (Pechlivani *et al*., 2022).

Mass spectrometry is a crucial tool for protein analysis. It precisely determines the exact molecular weight of the proteins, and also helps in analyzing proteolytic events and identifying chemical and post-translational modifications in proteins. Its high accuracy, speed and sensitivity make it ideal for ensuring protein purity, integrity and protein modifications (De Hoffmann *et al*., 2007; Witze *et al*., 2007). For example, Ke *et al*. identified a sphingomyelinase domain in apoB100 of electronegative Low-Density Lipoprotein (LDL) but not in native LDL. This domain is responsible for endothelial cell apoptosis (Ke *et al*., 2016).

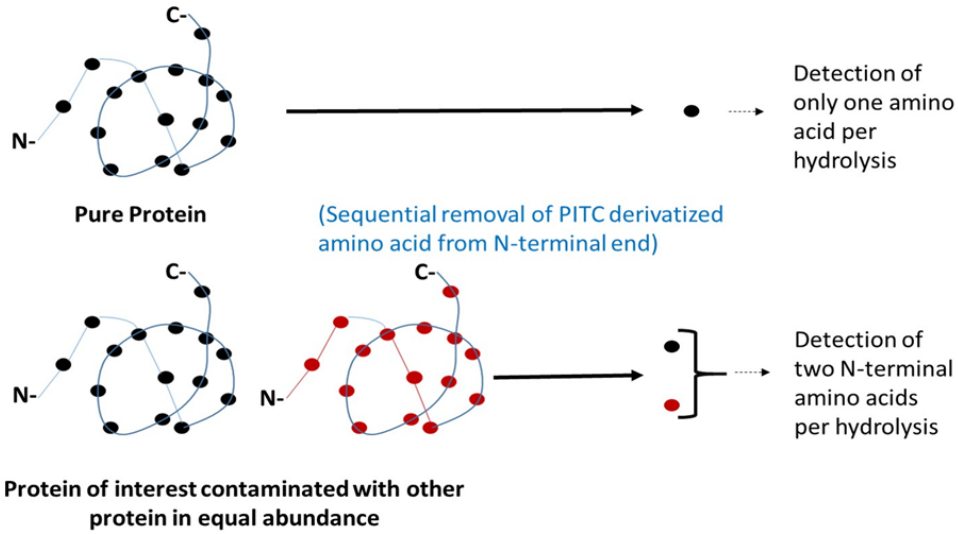

Fig. 2: Protein purity assessment by N-terminal analysis. Two amino acids are detected per hydrolysis if the protein of interest is contaminated with another protein

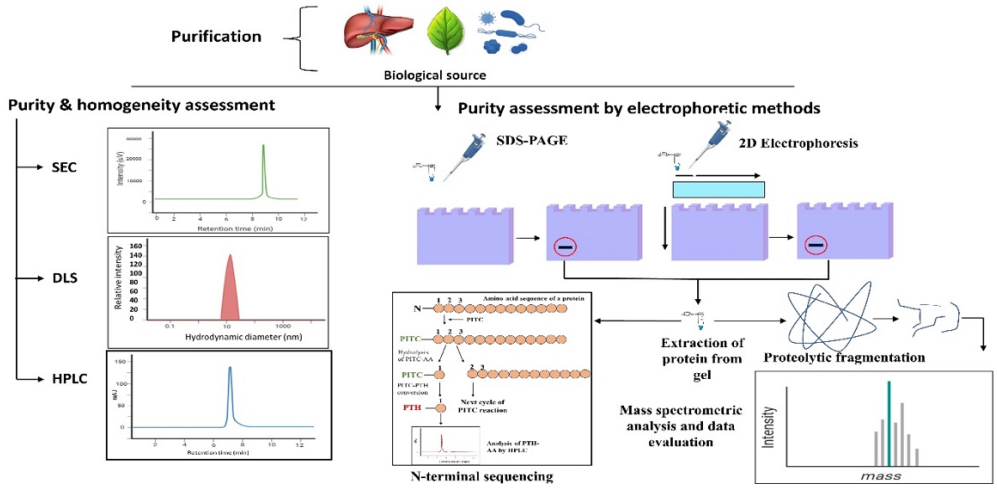

Fig. 3: Schematic representation of steps involved in protein purity assessment: The appearance of a single peak in SEC and HPLC, and mono dispersity of protein species in DLS indicates that the protein is pure. The appearance of a single band in SDS-PAGE and 2-DE, a single m/z ratio in mass spectral analysis, and one amino acid per cycle during N-terminal sequencing denote the presence of pure protein.

Both ESI and MALDI ionization efficiency is influenced by sample concentration and the complexity of the co-existing contaminants. Ultimately, irrespective of the ionization techniques, sample purity is vital for obtaining high-quality spectrum. The detection limit for both the methods is not just somewhere between picomoles-femtomoles, but even attomole limits can be detected (Martin *et al*., 2000; Valaskovic *et al*., 1996). The major drawback of most biological samples is protein contamination and dilution. Hence, several other techniques are usually coupled with mass spectrometry; the use of HPLC or CE, followed by lyophilization, can remove unwanted proteins/peptides and concentrate the protein of interest, respectively (Wahl *et al*., 1993). Tandem mass spectrometric methods like Collision-Induced Dissociation (CID) and infrared multiphoton dissociation are effective for analyzing smaller proteins, while Electron Capture Detectors (ECD) and electron transfer dissociation (ETD) are suited for larger protein analysis (Downard and Biemann, 1994; Syka *et al*., 2004; Tang *et al*., 1993; Zubarev *et al*., 1998). The protein sequence and certain covalent modifications can be determined by using proteolytic (trypsin) methods or chemicals (cyanogen bromide, formic acid, or hydroxylamine) that digest the protein into several peptides. The digested peptides are then matched against protein database with the help of protein identification programs such as MASCOT or SEQUEST. However, the protein of interest must be in the database (Eng *et al*., 1994; Liska and Shevchenko, 2003; Yates *et al*., 1996; Yates *et al*., 1995). If not, the method of choice may be to identify a short sequence of the unknown protein and synthesize oligonucleotide probes, which can then be used to determine the gene coding for that unknown protein. Finally, the gene must be cloned and sequenced. Mass spectrometry is an invaluable tool for identifying proteins from minute sample quantities. Furthermore, it enables the detection of protein mutations by analyzing the mass variation between the mutated and normal peptide, which is derived from proteolytic digestion of both proteins. For example, in an analysis of the mutation in human hemoglobin, MALDI mass spectroscopy showed the appearance of two peaks for the β-chain, and the mutated one was 14Da heavier than the normal one (Wada, 2002).

Purity of synthetic peptides and genetically engineered proteins are also assessed by mass spectrometry. A variety of modifications may occur during or after the synthesis of peptides. MALDI spectrum analysis can detect these errors and modifications. MS/MS analysis is the method of choice for detecting the contaminant proteins because of its ability to differentiate altered protein from normal protein by analyzing its mass and by peptide sequencing. ESI is particularly useful for studying the non-covalent interactions of proteins. In analyzing protein-protein, protein-ligand, or protein-drug interactions, the ESI spectra also provide information on the protein quaternary structure. For example, Loo *et al*. analysed native structure of the horse liver and yeast Alcohol Dehydrogenase (ADH) using ESI spectra; their findings clearly showed a dimeric form of ADH in horse liver and a tetrameric form in yeast (Loo, 1995). Recently, Reader *et al* assessed the monomeric purity of the therapeutic antibodies by ESI-MS (Reader *et al*., 2019). These

examples illustrate the sensitivity, rapidity, and importance of all the mass spectral methods in analyzing and characterizing proteins.

Other Analytical Techniques in Protein Purity Assessment

Most investigators employ various analytical techniques to assess protein purity and integrity. Size Exclusion Chromatography (SEC), Reverse-Phase High-Performance Liquid Chromatography (RP-HPLC), ultracentrifugation are important techniques.

SEC is the simplest method used in detecting impurities that differ in size. The concept of size-based separation of proteins was first introduced by Synge and Tiselius (Synge, 1950). SEC is used to detect molecular mass heterogeneity of proteins (Fig. 3). The gel-permeation chromatography, while less sensitive than gel electrophoresis, however, it could be highly accurate when the results are validated by an orthogonal method such as analytical ultracentrifugation. In addition, recent improvements in SEC columns have included a smaller particle size (≤ 2 μm), which have enhanced the resolution capacity of SEC columns. Thus, SEC may now be a better option in analyzing protein aggregates, impurities, and protein purity (Fekete *et al*., 2014; Hong *et al*., 2012). One major drawback of SEC would be, after elution, protein of interest gets diluted 10-20 times. Also, SEC would not be the better choice for eliminating the impurity if the contaminant and protein of interest have the same molecular weight.

Another analytical technique commonly used in the separation of proteins is RP-HPLC. Because of its simplicity and flexibility in coupling with other methods such as mass spectroscopy, RP-HPLC is a method of choice in analyzing protein purity (Fig.3). Most often C-18 columns are used, and proteins in turn elute by altering the polarity of the solvent system. Because a combination of various solvents is used in eluting the proteins, RP-HPLC can cause the irreversible denaturation of protein samples, thereby reducing the potential recovery of protein in its biologically active form. Thus, sometimes C-4 or C-8 columns are preferred because they require lower concentrations of organic solvents, and they have less affinity for proteins. Also, if the protein of interest differs in its hydrophobicity from that of contaminant protein, then only one could expect the separation of impurity from that of protein of interest. If both the proteins have similar hydrophobic properties, then there is a chance that the contaminant may coelute with the protein of interest. RP-HPLC is used to separate the tryptic digested peptides before mass spectroscopy analysis (Aguilar, 2004; Aguilar and Hearn, 1996; Frank *et al*., 1987). Recently, Guerin *et al* assessed the purity of recombinant proteins, CspZ and FhbA by HPLC (Guerin *et al*., 2024).

Analytical Ultracentrifugation (AUC) is a simple, precise, rapid, and non-destructive method. In contrast to other methods, after AUC analysis, samples can be recovered for further tests, as the purity analysis is performed in free solution in its native state. Sedimentation velocity and sedimentation equilibrium are the two different complementary views, which provide information on the macromolecule (protein) regarding its size-shape and molar masses, stoichiometries, association constants, and solution nonideality. Sedimentation velocity depends on the number of components in the sample that can be sedimented. If the protein of interest is the only component in the solution, then the appearance of a single, sharp boundary/peak would be expected (Fig. 3). If the difference in the molecular weight and size of the contaminant and the protein of interest is very small, then AUC is not a useful method for detecting impurity; the presence of single species in the solution needs to be confirmed by other methods (Cole *et al*., 2008; Pirie, 1940; Rhodes and Laue, 2009; Shedlovsky, 1943).

Light scattering techniques like Dynamic Light Scattering (DLS) and Static light Scattering (SLS) are used to assess protein aggregates (Nobbmann *et al*., 2007). Although DLS can detect small quantities of aggregates, the precise detection of each individual species is difficult. DLS is also not the method of choice in detecting close quaternary structures. SLS is a method that yields information regarding protein mass. As the light scattering signal is proportional to molecular mass, SLS is often coupled with SEC, HPLC and Asymmetric Flow–Field Flow Fractionation (AFFF) (Gabrielson *et al*., 2007; Sahin and Roberts, 2012). This helps in detecting solution homogeneity and in ensuring that the solution contains only one type of protein (Fig 3).

In addition to these methods used in assessing protein purity, few other properties of proteins may also be explored. For example, if the protein of our interest is an enzyme, obtaining constant specific activity over several purification steps is a reliable parameter of assessing protein purity. In addition, analyzing the amino acid composition of the purified protein and matching it with existing data can provide clues regarding purity. Immunodetection of proteins by specific antibodies is useful in assessing protein purity but cannot be used to rule out the presence of contaminants. The availability of all these techniques makes protein purity assessment easier and allows for cross checking purity by one or another method.

Conclusion

Analyzing the purity of test proteins is an essential requirement in biological research. However, homogeneous protein purification from any biological sample is a daunting task. Similarly, assessing the purity of a purified protein is challenging and almost always involves a combination of several techniques. The

following observations can be used to establish the purity of a protein:

- Visualization of a single band on SDS-PAGE (more precisely on 2-DE after silver staining).
- Similarity in N-terminal sequence of the test protein to a known sequence (if the study protein is a known one).
- Presence of a single m/z ratio.
- Constant specific activity over several purification steps, if the study protein is an enzyme.
- If the protein under study is known, its amino acid composition should match the known protein.

Acknowledgment

The authors acknowledge the Institute of Excellence, University of Mysore, Council of Scientific and Industrial Research, New Delhi, India. The authors thank the editorial assistance of Rebecca Bartow, PhD, Department of Scientific Publications, Texas Heart Institute, Houston, Texas.

Funding information

This study received no financial support.

Conflicts of Interest

The authors of this manuscript declare no conflict of interest.

Authors' contributions

Shiva Siddappa: Writing Original Draft.
Liang-Yin Ke, Chu-Huang Chen: Writing Review and Editing.
Gopal Kedihithlu Marathe: Conceptualization, Writing Review and Editing.

Ethics

Authors should address any ethical issues that may arise after the publication of this manuscript.

References

Aguilar, M. I. (2004). Reversed-phase high-performance liquid chromatography. *Methods in Molecular Biology*, *251*, 9–22. https://doi.org/10.1385/1-59259-742-4

Aguilar, M.-I., & Hearn, M. T. (1996). High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins. *Methods in Enzymology*, *270*, 3–26. https://doi.org/10.1016/s0076-6879(96)70003-4

Berrow, N., de Marco, A., Lebendiker, M., Garcia-Alai, M., Knauer, S. H., Lopez-Mendez, B., Matagne, A., Parret, A., Remans, K., Uebel, S., & Raynal, B. (2021). Quality control of purified proteins to improve data quality and reproducibility: results from a large-scale survey. *European Biophysics Journal*, *50*(3–4), 453–460. https://doi.org/10.1007/s00249-021-01528-2

Chiari, M., Righetti, P. G., Negri, A., Ceciliani, F., & Ronchi, S. (1992). Preincubation with cysteine prevents modification of sulfhydryl groups in proteins by unreacted acrylamide in a gel. *ELECTROPHORESIS*, *13*(1), 882–884. https://doi.org/10.1002/elps.11501301193

Cole, J. L., Lary, J. W., P. Moody, T., & Laue, T. M. (2008). Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods in Cell Biology*, *84*, 143–179. https://doi.org/10.1016/s0091-679x(07)84006-4

De Hoffmann, E., & Stroobant, V. (2007). *Mass spectrometry: principles and applications*.

De Marco, A., Berrow, N., Lebendiker, M., Garcia-Alai, M., Knauer, S. H., Lopez-Mendez, B., Matagne, A., Parret, A., Remans, K., Uebel, S., & Raynal, B. (2021). Quality control of protein reagents for the improvement of research data reproducibility. *Nature Communications*, *12*(1), 2795. https://doi.org/10.1038/s41467-021-23167-z

De Souza, N. (2013). Studying function at the genome scale. *Nature Methods*, *10*(10), 937–937. https://doi.org/10.1038/nmeth.2679

Downard, K. M., & Biemann, K. (1994). The effect of charge state and the localization of charge on the collision-induced dissociation of peptide ions. *Journal of the American Society for Mass Spectrometry*, *5*(11), 966–975. https://doi.org/10.1016/1044-0305(94)80015-4

Edman, P. (1950). Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chemica Scandinavica*, *4*, 283–293.

Eng, J. K., McCormack, A. L., & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, *5*(11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2

Fekete, S., Beck, A., Veuthey, J.-L., & Guillarme, D. (2014). Theory and practice of size exclusion chromatography for the analysis of protein aggregates. *Journal of Pharmaceutical and Biomedical Analysis*, *101*, 161–173. https://doi.org/10.1016/j.jpba.2014.04.011

Fowler, E., Moyer, M., Krishna, R. G., Chin, C. C. Q., & Wold, F. (1996). Removal of N-Terminal Blocking Groups from Proteins. *Current Protocols in Protein Science*, *3*(1), 7–11. https://doi.org/10.1002/0471140864.ps1107s03

Frank, J., Braat, A., & Duine, J. A. (1987). Assessment of protein purity by chromatography and multiwavelength detection. *Analytical Biochemistry*, *162*(1), 65–73. https://doi.org/10.1016/0003-2697(87)90010-8

Gabrielson, J. P., Brader, M. L., Pekar, A. H., Mathis, K. B., Winter, G., Carpenter, J. F., & Randolph, T. W. (2007). Quantitation of Aggregate Levels in a Recombinant Humanized Monoclonal Antibody Formulation by Size-Exclusion Chromatography, Asymmetrical Flow Field Flow Fractionation, and Sedimentation Velocity. *Journal of Pharmaceutical Sciences*, *96*(2), 268–279. https://doi.org/10.1002/jps.20760

Görg, A., Postel, W., & Günther, S. (1988). Two-dimensional electrophoresis. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, *9*(9), 531–546. https://doi.org/10.1002/elps.1150090913

Guérin, M., Vandevenne, M., Brans, A., Matagne, A., Marquant, R., Prost, E., Octave, S., Avalle, B., Maffucci, I., & Padiolleau-Lefèvre, S. (2024). Production, purification, and quality assessment of borrelial proteins CspZ from Borrelia burgdorferi and FhbA from Borrelia hermsii. *Applied Microbiology and Biotechnology*, *108*(1), 425. https://doi.org/10.1007/s00253-024-13195-2

Guo, A., Han, M., Martinez, T., Ketchem, R. R., Novick, S., Jochheim, C., & Balland, A. (2008). Electrophoretic evidence for the presence of structural isoforms specific for the IgG2 isotype. *Electrophoresis*, *29*(12), 2550–2556. https://doi.org/10.1002/elps.200800083

Hong, P., Koza, S., & Bouvier, E. S. P. (2012). A Review Size-exclusion Chromatography for the Analysis of Protein Biotherapeutics and Their Aggregates. *Journal of Liquid Chromatography & Related Technologies*, *35*(20), 2923–2950. https://doi.org/10.1080/10826076.2012.743724

Ke, L.-Y., Chan, H.-C., Chen, C.-C., Lu, J., Marathe, G. K., Chu, C.-S., Chan, H.-C., Wang, C.-Y., Tung, Y.-C., McIntyre, T. M., Yen, J.-H., & Chen, C.-H. (2016). Enhanced Sphingomyelinase Activity Contributes to the Apoptotic Capacity of Electronegative Low-Density Lipoprotein. *Journal of Medicinal Chemistry*, *59*(3), 1032–1040. https://doi.org/10.1021/acs.jmedchem.5b01534

Kornberg, A. (2003). Ten commandments of enzymology, amended. *Trends in Biochemical Sciences*, *28*(10), 515–517. https://doi.org/10.1016/j.tibs.2003.08.007

Kornberg, A. (2009). Why purify enzymes? *Methods in Enzymology*, *463*, 3–6. https://doi.org/10.1016/S0076-6879(09)63001-9

Kumar, R., Guttman, A., & Rathore, A. S. (2022). Applications of capillary electrophoresis for biopharmaceutical product characterization. *Electrophoresis*, *43*(2), 143–166. https://doi.org/10.1002/elps.202100182

Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. In *Nature* (Vol. 227, Issue 5259, pp. 680–685). https://doi.org/10.1038/227680a0

Liska, A. J., & Shevchenko, A. (2003). Combining mass spectrometry with database interrogation strategies in proteomics. *Trends in Analytical Chemistry*, *22*(5), 291–298. https://doi.org/10.1016/s0165-9936(03)00507-7

Loo, J. A. (1995). Observation of large subunit protein complexes by electrospray ionization mass spectrometry. *Journal of Mass Spectrometry*, *30*(1), 180–183. https://doi.org/10.1002/jms.1190300127

Magdeldin, S., Enany, S., Yoshida, Y., Xu, B., Zhang, Y., Zureena, Z., Lokamani, I., Yaoita, E., & Yamamoto, T. (2014). Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clinical Proteomics*, *11*(1), 16. https://doi.org/10.1186/1559-0275-11-16

Marathe, G. K., Zimmerman, G. A., & McIntyre, T. M. (2003). Platelet-activating Factor Acetylhydrolase, and Not Paraoxonase-1, Is the Oxidized Phospholipid Hydrolase of High Density Lipoprotein Particles. *Journal of Biological Chemistry*, *278*(6), 3937–3947. https://doi.org/10.1074/jbc.m211126200

Martin, S. E., Shabanowitz, J., Hunt, D. F., & Marto, J. A. (2000). Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry*, *72*(18), 4266–4274. https://doi.org/10.1021/ac000497v

Mohan, S. B. (1992). Determination of purity and yield. *Methods in Molecular Biology*, *11*, 307–323. https://doi.org/10.1385/0-89603-213-2:307

Nebija, D., Urban, E., Stessl, M., Noe, C. R., & Lachmann, B. (2011). 2-DE and MALDI-TOF-MS analysis of therapeutic fusion protein abatacept. *Electrophoresis*, *32*(12), 1438–1443. https://doi.org/10.1002/elps.201100029

Nobbmann, U., Connah, M., Fish, B., Varley, P., Gee, C., Mulot, S., Chen, J., Zhou, L., Lu, Y., Sheng, F., Yi, J., & Harding, S. E. (2007). Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnology and Genetic Engineering Reviews*, *24*(1), 117–128. https://doi.org/10.1080/02648725.2007.10648095

O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, *250*(10), 4007–4021. https://doi.org/10.1016/s0021-9258(19)41496-8

Öztürk, C., & Küfrevioğlu, Ö. İ. (2024). Affinity gel synthesis from the p-aminobenzoic acid derivative 4-amino-2-methylbenzoic acid and purification of polyphenol oxidase from various plant sources. *Protein Expression and Purification*, *219*, 106474. https://doi.org/10.1016/j.pep.2024.106474

Pechlivani, N., Kearney, K. J., Tiede, C., Cheah, R., Phoenix, F., Ponnambalam, S., Ault, J. R., McPherson, M. J., Tomlinson, D. C., & Ajjan, R. A. (2022). Affinity purification of fibrinogen using an Affimer column. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1866*(5), 130115. https://doi.org/10.1016/j.bbagen.2022.130115

Pirie, N. W. (1940). The Criteria of Purity Used in the Study of Large Molecules of Biological Origin. *Biological Reviews*, *15*(4), 377–404. https://doi.org/10.1111/j.1469-185x.1940.tb00945.x

Rabilloud, T., Chevallet, M., Luche, S., & Lelong, C. (2010). Two-dimensional gel electrophoresis in proteomics: Past, present and future. *Journal of Proteomics*, *73*(11), 2064–2077. https://doi.org/10.1016/j.jprot.2010.05.016

Rajabi, N., Safarnejad, M. R., Rakhshandehroo, F., Shamsbakhsh, M., & Rabbani, H. (2022). *Developing of specific monoclonal recombinant antibody fused to alkaline phosphatase (AP) for one-step detection of fig mosaic virus*. *12*, 88. https://doi.org/10.1007/s13205-022-03154-x

Raynal, B., Lenormand, P., Baron, B., Hoos, S., & England, P. (2014). Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories*, *13*(1), 180. https://doi.org/10.1186/s12934-014-0180-6

Reader, P. P., Olkhov, R. V., Reeksting, S., Lubben, A., Hyde, C. J., & Shaw, A. M. (2019). A rapid and quantitative technique for assessing IgG monomeric purity, calibrated with the NISTmAb reference material. *Analytical and Bioanalytical Chemistry*, *411*(24), 6487–6496. https://doi.org/10.1007/s00216-019-02029-0

Rhodes, D. G., & Laue, T. M. (2009). Chapter 38 Determination of Protein Purity. *Methods in Enzymology*, *463*, 677–689. https://doi.org/10.1016/s0076-6879(09)63038-x

Rodrigo, L., Mackness, B., Durrington, P. N., Hernandez, A., & Mackness, M. I. (2001). Hydrolysis of platelet-activating factor by human serum paraoxonase. *Biochemical Journal*, *354*(1), 1–7. https://doi.org/10.1042/bj3540001

Sahin, E., & Roberts, C. J. (2012). Size-Exclusion Chromatography with Multi-angle Light Scattering for Elucidating Protein Aggregation Mechanisms. *Aggregation of Therapeutic Proteins*, *899*, 403–423. https://doi.org/10.1007/978-1-61779-921-1_25

Salarifar, A., Khalili, S., & Rasaee, M. J. (2023). Recombinant highly antigenic truncated fusion-based protein as a diagnostic antigen for anti-SARS-CoV-2 nucleocapsid antibody ELISA. *Protein Expression and Purification*, *203*, 106200. https://doi.org/10.1016/j.pep.2022.106200

Shedlovsky, T. (1943). Criteria of Purity of Proteins. *Annals of the New York Academy of Sciences*, *43*(6), 259–272. https://doi.org/10.1111/j.1749-6632.1943.tb54083.x

Slibinskas, R., Ražanskas, R., Zinkevičiūtė, R., & Čiplys, E. (2013). Comparison of first dimension IPG and NEPHGE techniques in two-dimensional gel electrophoresis experiment with cytosolic unfolded protein response in Saccharomyces cerevisiae. *Proteome Science*, *11*(1), 36. https://doi.org/10.1186/1477-5956-11-36

Speicher, K. D., Gorman, N., & Speicher, D. W. (2009). N-Terminal Sequence Analysis of Proteins and Peptides. *Current Protocols in Protein Science*, *57*(1), 10–11. https://doi.org/10.1002/0471140864.ps1110s57

Stern, A. S., Podlaski, F. J., Hulmes, J. D., Pan, Y. C., Quinn, P. M., Wolitzky, A. G., Familletti, P. C., Stremlo, D. L., Truitt, T., & Chizzonite, R. (1990). Purification to homogeneity and partial characterization of cytotoxic lymphocyte maturation factor from human B-lymphoblastoid cells. *Proceedings of the National Academy of Sciences*, *87*(17), 6808–6812. https://doi.org/10.1073/pnas.87.17.6808

Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., & Hunt, D. F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences*, *101*(26), 9528–9533. https://doi.org/10.1073/pnas.0402700101

Synge, R. L. M. (1950). Fractionation of hydrolysis products of amylose by electrokinetic ultrafiltration in an agaragar jelly. *Biochemical Journal*, *2*, 41–42.

Tan, S. Y., & Berman, Y. E. (2008). Robert Koch (1843-1910): father of microbiology and Nobel laureate. *Singapore Medical Journal*, *49*(11), 854–855.

Tang, X. Jun., Thibault, Pierre., & Boyd, R. K. (1993). Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Analytical Chemistry*, *65*(20), 2824–2834. https://doi.org/10.1021/ac00068a020

Tipton, J. D., Tran, J. C., Catherman, A. D., Ahlf, D. R., Durbin, K. R., & Kelleher, N. L. (2011). Analysis of Intact Protein Isoforms by Mass Spectrometry. *Journal of Biological Chemistry*, *286*(29), 25451–25458. https://doi.org/10.1074/jbc.r111.239442

Valaskovic, G. A., Kelleher, N. L., & McLafferty, F. W. (1996). Attomole Protein Characterization by Capillary Electrophoresis-Mass Spectrometry. *Science*, *273*(5279), 1199–1202. https://doi.org/10.1126/science.273.5279.1199

Wada, Y. (2002). Advanced analytical methods for hemoglobin variants. *Journal of Chromatography B*, *781*(1–2), 291–301. https://doi.org/10.1016/s1570-0232(02)00558-5

Wahl, J. H., Goodlett, D. R., Udseth, H. R., & Smith, R. D. (1993). Use of small-diameter capillaries for increasing peptide and protein detection sensitivity in capillary electrophoresis-mass spectrometry. *Electrophoresis*, *14*(1), 448–457. https://doi.org/10.1002/elps.1150140170

Watson, A. D., Berliner, J. A., Hama, S. Y., La Du, B. N., Faull, K. F., Fogelman, A. M., & Navab, M. (1995). Protective effect of high density lipoprotein associated paraoxonase. Inhibition of the biological activity of minimally oxidized low density lipoprotein. *Journal of Clinical Investigation*, *96*(6), 2882–2891. https://doi.org/10.1172/jci118359

Wellner, D., Panneerselvam, C., & Horecker, B. L. (1990). Sequencing of peptides and proteins with blocked N-terminal amino acids: N-acetylserine or N-acetylthreonine. *Proceedings of the National Academy of Sciences*, *87*(5), 1947–1949. https://doi.org/10.1073/pnas.87.5.1947

Westermarck, J., Ivaska, J., & Corthals, G. L. (2013). Identification of Protein Interactions Involved in Cellular Signaling. *Molecular & Cellular Proteomics*, *12*(7), 1752–1763. https://doi.org/10.1074/mcp.r113.027771

Whitmore, C. D., & Gennaro, L. A. (2012). Capillary electrophoresis-mass spectrometry methods for tryptic peptide mapping of therapeutic antibodies. *Electrophoresis*, *33*(11), 1550–1556. https://doi.org/10.1002/elps.201200066

Witze, E. S., Old, W. M., Resing, K. A., & Ahn, N. G. (2007). Mapping protein post-translational modifications with mass spectrometry. *Nature Methods*, *4*(10), 798–806 https://doi.org/10.1038/nmeth1100

Yates, J. R., Eng, J. K., Clauser, K. R., & Burlingame, A. L. (1996). Search of sequence databases with uninterpreted high-energy collision-induced dissociation spectra of peptides. *Journal of the American Society for Mass Spectrometry*, *7*(11), 1089–1098. https://doi.org/10.1016/s1044-0305(96)00079-7

Yates, J. R., Eng, J. K., McCormack, A. L., & Schieltz, David. (1995). Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Analytical Chemistry*, *67*(8), 1426–1436. https://doi.org/10.1021/ac00104a020

Zannis, V., Doyle, D., & Martin, D. W. (1978). Purification and characterization of human erythrocyte purine nucleoside phosphorylase and its subunits. *Journal of Biological Chemistry*, *253*(2), 504–510. https://doi.org/10.1016/s0021-9258(17)38238-8

Zhao, S. S., & Chen, D. D. Y. (2014). Applications of capillary electrophoresis in characterizing recombinant protein therapeutics. *Electrophoresis*, *35*(1), 96–108. https://doi.org/10.1002/elps.201300372

Zhu, Z., Lu, J. J., & Liu, S. (2012). Protein separation by capillary gel electrophoresis: A review. *Analytica Chimica Acta*, *709*, 21–31. https://doi.org/10.1016/j.aca.2011.10.022

Zubarev, R. A., Kelleher, N. L., & McLafferty, F. W. (1998). Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *Journal of the American Chemical Society*, *120*(13), 3265–3266. https://doi.org/10.1021/ja973478k