

Electrophoresis

I complete the series of lectures on protein purification by talking about electrophoresis - which can be both a preparative and an analytical method. Scopes' review, Protein purification in the nineties, stresses that the action is at the two ends of the size scale, large amounts of purified protein for industrial uses, very small amounts for sequencing by automated Edman degradation, around 1 μg , and even less for mass spectrometric analysis, 1-100 ng. Electrophoresis techniques devised for analysis can be the final step of preparation for these methods. The term "preparative electrophoresis" refers to procedures able to purify at least a few mg of protein at a time, as distinguished from analytical electrophoresis at the microgram level.

By electrophoresis we mean any method which separates proteins by moving them in an electric field. This topic can be subdivided in many ways: by whether it is used preparatively or analytically; by the type of electrophoresis, especially whether it is a rate method or an end point method; by the material used as support during electrophoresis; by whether the protein is expected to remain biologically active after electrophoresis; and by methods of analysis of the results. However, one case among all these is best known; as Scopes says, many people will use polyacrylamide gel electrophoresis in presence of sodium dodecyl sulfate without being aware that there are any other procedures!

I shall talk about the material used as support first, because one choice is so dominant, although there have been many others. The first electrophoresis was Tiselius free-boundary electrophoresis, of rather concentrated protein solutions in a large U-shaped tube with buffer overlaid, so that the movement of the proteins could be observed as movement of boundaries of refractive index zones. The change in refractive index, called schlieren, was plotted as the first derivative, so that a boundary of concentration changing sigmoidally came out as a peak. But such electrophoresis is difficult because of the difficulty of avoiding mixing by convection, and it was replaced in the 1950s by solid, hydrophilic supports which prevented convection and diminished diffusion. At first this was filter paper or cellulose acetate, but these have limited capacity and resolution. Then starch gel was introduced, which has large pores which do not separate proteins by size, and hold a lot of protein; agarose is used the same way. Starch and agarose gels are still used for one main purpose, for which cellulose acetate was previously used: to separate **isozymes**, or more broadly **isoforms** of a protein - forms which have the same or similar biological activity but differ in some way which makes them physically separable, usually charge which makes them separable by electrophoresis. [Usually we mean related forms of a protein which have the same molecular weight (the most famous are the isoforms of lactate dehydrogenase, five forms made up of different numbers of two types of subunit). Therefore there is no point trying to separate by molecular weight, one wants to separate by charge alone, and therefore uses a large pore gel such as starch or agarose which doesn't try to separate by size. It should be noted that isoforms are defined operationally, as having activity in the same assay but being separable by some method, usually electrophoresis; the term says nothing about how related they are. At one extreme, they may be unrelated proteins which happen to be active in the same assay, particularly if the assay is non-specific such as proteolysis. They may represent duplications of the same gene, with related but non-identical sequences, such as fetal and adult hemoglobin, or the lactate dehydrogenase isozymes. They may be allelic variants of the same protein, products of wild-type and mutant forms of the same gene, as normal and sickle-cell hemoglobin. They may represent natural different forms of the same protein resulting from differences in post-translational modification, such as pancreatic ribonuclease A, which isn't glycosylated, and B, which is. They may represent artifactually different forms of the same protein, for instance resulting from proteolysis or deamidation during purification, or protein with and without some bound ligand. Scopes notes that in isoelectric focusing a protein may bind different carrier ampholyte molecules, resulting in apparent different bands of different pI. The point is to realize that there are

many possibilities; too many people assume that electrophoretically separable isozymes are the phenotypes of different alleles of the same gene, without investigating further.]

One other "support material" should be mentioned: sucrose density gradients, which minimize convection but are free flowing liquids, and can be drained carefully from a column after electrophoretic separation has taken place, with no difficulty in recovering the protein. They are therefore useful in preparative scale electrophoresis.

But the paramount material is of course polyacrylamide, a hydrophobic gel prepared by polymerizing acrylamide solutions with some methylenebisacrylamide as cross-linker, usually using ammonium persulfate as initiator and tetramethylethylene diamine - TEMED - as accelerator (**draw reaction**). These gels have pores of the size of protein molecules, so that proteins being electrophoresed through them are separated on a size basis - large proteins move with more difficulty than small proteins - whether or not they are also being separated on the basis of their charge. Gels are described by what % acrylamide they contain, from 12% acrylamide for small proteins, 10 to 20,000, to 6% for large proteins over 100,000; that is about as low a percentage as can be worked with, though I have seen a paper title about using 2% acrylamide gels strengthened by agarose to characterize very large proteins. Hydroxymethyl methacrylate is one alternative being pushed by a company. Sometimes granular polyacrylamide, or even Sephadex or other carbohydrate polymer, is used as a supporting material for electrophoresis in a column or on a flat bed, whether or not the proteins enter the pores, because after electrophoretic separation proteins can readily be washed out, whereas proteins in a solid polyacrylamide gel are not so easily removed - usually they have to be electroeluted out. There are preparative electrophoresis cells which have a stream of buffer passing across the bottom of a column of gel, so that proteins emerging from the gel by electrophoresis are washed aside into a fraction collector; however, there tends to be a lot of dilution. Scopes says that while there are preparative electrophoresis cells and they do work for some, their use is not to be undertaken lightly; I think it depends on having someone in the lab who will make it a major objective of life to make the system work, and when that person leaves the preparative electrophoresis system will stop being used.

[It should be remembered that there is likely to be some unpolymerized, free acrylamide in a gel, which may react with cysteine residues in proteins, or even with other amino acids. Remaining ammonium persulfate may also oxidize cysteines and methionines. These reactions may affect biological activity, if this is preparative scale electrophoresis, and may also affect sequencing of the protein or peptides by mass spectrometry, but are not much of a problem for usual analytical gel electrophoresis.]

Electrophoretic methods may be divided broadly into rate and equilibrium methods. Rate methods include the usual electrophoresis, in which different molecules migrate through a gel or other support at different rates and are distinguished by how far they have gone when the current is shut off, and isotachopheresis, a method related to isoelectric focusing in which proteins are focused into moving bands separated by small molecules of intermediate pIs; I don't know much about it as a method. Equilibrium methods include gradient gels, about which more later, and isoelectric focusing, which may be used preparatively, usually in a sucrose density gradient in a column, or analytically in a gel. The Pharmacia PhastSystem uses rapid isoelectric focusing in small gels.

Isoelectric focusing involves electrophoretically moving a protein in a pH gradient until it reaches the pH at which it is isoelectric; it then stops. The pH gradient consists of a variety of molecules, peptides and small synthetic proteins, with different isoelectric points, called **ampholytes**. Initially these are dispersed throughout the gradient, but when an electric field is applied, these molecules, and the proteins present, migrate toward one or the other electrode. The positive electrode will be in a fairly strong base solution such as ethylenediamine, the negative electrode in a fairly strong acid such as phosphoric. Eventually the ampholytes and the protein reach pH values at which they are isoelectric, and stop. One can buy - they are expensive - wide range mixtures of ampholytes, generating

a gradient from pH 3 to 10, or narrow ranges covering only 2 pH units. With these latter one can separate proteins differing in pI by as little as 0.02 pH unit. One uses marker proteins of known pI to calibrate the gel.

As a preparative technique, in a sucrose gradient, isoelectric focusing has some constraints - the protein of interest must be stable, not denatured, at its pI, and mustn't precipitate out, such a precipitate will fall through the sucrose gradient and mess it up as well as not remaining at its pI; indeed even some other protein precipitating can ruin the gradient. Precipitation is less of a problem in analytical isoelectric focusing at low concentration, and it may not be necessary that the protein retain its activity. As an analytical technique it is good, both to show whether the protein is pure and to give you the pI, which may be useful in improving the purification; but note that small differences such as post-translational modifications or different numbers of buffer molecules binding to the protein may give you multiple bands of the same protein.

Isoelectric focusing, in a small tube of gel, is now widely employed as a first dimension separation in two dimensional electrophoretic analytical separation, 2D gels for short, typically used on crude extracts to see whether there have been changes in proteins expressed under different conditions of a tissue. After focusing, the gel is extruded from the tube and laid along the top of a slab gel of polyacrylamide containing sodium dodecyl sulfate; it may be heated in 1% SDS to assure denaturation of proteins and binding of SDS. The proteins are then electrophoresed down into the slab gel, and visualized by Coomassie Blue or silver staining, or autoradiography to locate a particular labeled protein. Because they have been separated in two dimensions as many as 1000 different proteins can be seen on such a gel, and it becomes necessary to use a computer to analyze the results if one is comparing samples obtained under different conditions to see what changes have occurred, the more so since the runs are never exactly reproducible. However, if all you are interested in is sequence, and you know which spot you want, you can transfer the proteins to a polyvinylidene difluoride membrane and sequence directly on the membrane. Or the protein spot may be digested by a protease, either in the gel or after transfer to a membrane, and the peptides eluted, separated by hplc, and sequenced, either by Edman degradation or with even greater sensitivity by mass spectrometry. I refer you to a paper by Scott Paterson, *Anal. Biochem.* **221**:1-15 (1994) on this.

The commonest electrophoresis is in a polyacrylamide gel for analytical purposes, to determine whether the purified protein is homogeneous and to determine subunit molecular weight, or in some procedures native molecular weight. It can also be used before sequencing, as just described for spots in 2D. Some workers still also use analytical ultracentrifugation to assess purity. Another technique, used at a particularly small scale, and more for small molecules than for proteins, is capillary electrophoresis, about which I know little. Automated DNA sequencing now uses electrophoresis in capillaries rather than in polyacrylamide gels.

Two general things should be remembered: first, it is never possible to prove that a proteins is really homogeneous, only that impurities have not been observed by the techniques employed. The most troublesome impurity is an inactive form of the same protein, inseparable but affecting quantitative measurement of the properties of the protein. But it wouldn't matter for sequencing, unless its difference was that it had lost N-terminal sequence. Second, the technique used may show several active forms of the same protein - real or artifactual isoforms - whose differences may or may not be significant. Different forms with different kinetic constants can mess up kinetic analysis of the reaction. Sometimes careful electrophoresis has shown that a protein believed to be a homotetramer - four identical subunits - really was made up of four slightly different subunits.

Analytical gel electrophoresis includes isoelectric focusing - which I have already mentioned - 'native' gel electrophoresis, gradient gel electrophoresis and SDS and other denaturing detergent electrophoresis. In the last three the size of the protein, its movement through the variously sized pores of the gel, is an important factor, in gradient and SDS gel electrophoresis the *only* factor.

'Native' gel electrophoresis observes the protein in its native state, perhaps containing multiple subunits, in a non-denaturing buffer. In this situation the protein is moved by the interaction of its side chain charges with the electrical field, and how fast it moved is determined by its charge density - the net charge divided by the mass of the protein - as well as by finding its way through the pores. A large but highly charged protein may move the same as a smaller, less charged protein. Consequently one cannot determine the molecular weight from a single native gel at a single acrylamide concentration. However, if electrophoresis is carried out on a number of gels, with the same buffer conditions but varying acrylamide concentration, the procedure of Hedrick and Smith, the charge effect is factored out and the difference in movement on the various gels is due entirely to the size of the molecule. A plot of $\log R_m$ (movement relative to the tracking dye) vs % acrylamide should be linear, and the slope of such a line is proportional to the molecular weight of the protein. Thus with electrophoresis of several standards - bovine serum albumin, which contains some dimer and trimer, is useful - on gels at three or four acrylamide concentrations a standard curve can be set up, and the slope of $\log R_m$ vs. % acrylamide for an unknown protein plotted on this curve and its native molecular weight found. This is especially useful if the protein can be found by its activity after the run.

A gradient gel is a vertical slab gel with a gradient of acrylamide concentration, increasing down the gel, so that the pores get smaller and smaller. Proteins migrate until they can no longer find a pore to move through. This is thus an equilibrium method rather than a rate method, and is run longer. The final position depends on the molecular weight, a plot of R_m vs. \log mol. wt. is linear, and standards can be run on the same gel at the edge. You can pour them yourself, but they are much more reproducible if bought from Pharmacia. One good point about gradient gels is that the bands get sharper as they slow down, so that quite small differences in mol. wt. can be seen.

For both native and gradient gels it should be remembered that perhaps one-fifth of proteins are **cationic** at neutral pH, and will electrophorese in the other direction. This is where knowledge of the pI is important. One can of course make lower pH gels for cationic proteins and run them with electrodes reversed, but it may be hard to get mol. weight standards for gradient gels.

SDS - sodium dodecyl sulfate, unless you were a radical in the sixties - is a powerful detergent, which usually will denature completely a protein without disulfide bonds; heating in 1% mercaptoethanol is usually included in the preparation of the sample to reduce any disulfide bonds. Proteins usually are not expected to renature after SDS gel electrophoresis, but if they have not been heated too much, have not dissociated into unlike subunits or lost a cofactor, and especially if pure dodecyl sulfate is used rather than the commercial mixture containing C₁₀, C₁₂ and C₁₄ alkylsulfonates, they often can be renatured on washing out the SDS. Also, antibodies usually still recognize the denatured protein if they can get at it; hence the 'western blot', transfer of proteins to a membrane after electrophoresis and exposure to specific antibodies, followed by exposure to a second, enzyme-linked antibody which visualizes the specific band wanted. There is also a procedure using the **cationic** detergent cetyl trimethylammonium bromide, which is claimed to allow determination of molecular weight without complete denaturation. In my experience the bands are not sharp.

The point of SDS gel electrophoresis is not only that the protein is denatured and moves as random chains of the monomer; SDS associates with the peptide chain, about one per two amino acids, so that the protein-SDS complex has a large negative charge from the SDS, drowning out the effect of charges on the side chains. Thus all proteins have essentially the same charge density, and movement in gel electrophoresis is essentially a function only of the size of the protein, and plots of R_m vs \log mol. wt. are linear, though one should be cautious at very low mol. wt., below 12,000. A Tris-Tricine buffer is better than the traditional Tris-glycine buffer here. And if the protein is heavily glycosylated its apparent mol. wt. will be off, as the carbohydrate does not adsorb SDS and the charge density will be lower.

SDS has one drawback, that gels must be run at room temperature, because below 23° C SDS forms large micelles and precipitates out. Of course if the protein is denatured anyway this isn't a problem; but you can run gels in the cold room if you use **lithium** dodecyl sulfate - Dr. Niederman's lab does this, to keep chromophores associated with the protein.

Both native and SDS gels are usually run as 'discontinuous' systems, with a 'stacking gel' above the main separating gel, in which the proteins are concentrated into narrow bands. The stacking gel has a very low acrylamide concentration, 2.5% to 4%, to minimize diffusion and convection without any size separation. The important feature is that its pH is 1.5 to 2 units nearer neutrality than the running gel (lower if the proteins electrophorese as anions, higher if they are cations). The other important feature is that the reservoir buffer uses an ion - the anion if the proteins run as anions - which is partly charged at the running gel pH, but only very slightly charged at the stacking gel pH, so that it has a net mobility **less** than the proteins. For instance, is the most used basic gel system, the running gel is Tris Cl at pH 8.8, the stacking gel Tris Cl or phosphate at pH 7, while the upper reservoir buffer is Tris glycine at pH 8.3. Glycine has a basic pK_2 9.87, so it is only about 1/1000th anion at pH 6.87, its mobility is lower than that of the proteins. The proteins thus migrate behind the high mobility chloride or phosphate ion of the stacking gel, ahead of the glycine, but in order to carry enough current they **concentrate** into very narrow bands, much narrower and more concentrated than the samples applied. The mathematics of this was worked out by Kohlrausch in 1897. Once the proteins get into the running gel they slow down, because of the smaller pores, and the glycine speeds up, because the pH is higher; the tracking dye added with the proteins remains at the front between the glycine and the chloride. The proteins are now no longer stacked, and begin to broaden their bands again by diffusion; but they also separate, on the basis of their mobilities, size and net charge in native gel electrophoresis or size alone in SDS gels.

Proteins generally are located by some general staining technique, often after washing the gel with a methanol-acetic acid mixture which fixes the proteins so that they don't wash out, but removes SDS. Isoelectric focusing gels have to be washed thoroughly to remove the ampholyte. The most often used stain is Coomassie Blue G-250 in a methanol-acetic acid mixture, even though the gel then has to be washed further to remove excess stain. The literature is full of variations on this, but I've tried a lot of them and they are generally less sensitive. An older stain, not much used now, was Buffalo Black, also known as Amido Schwarz. The next most often used is silver staining, which is very sensitive but requires care and very good water not to get a dark background. Rosenberg also gives a procedure for staining with Cu^{++} or Zn^{++} , which gives clear bands against a cloudy background but is supposed to be very sensitive. I haven't used it. Another dye, especially on blots, is 0.1% Ponceau S; yet another is Nile Red. The newest thing is fluorescent stains, SYPRO Red and Orange, which give fluorescent red or orange bands and are supposed to be as sensitive as silver staining but easier (the stain is preferably in 7.5% acetic acid). They actually bind to the SDS associated with the protein, and thus have their fluorescence enhanced in the hydrophobic environment; the method is thus limited to SDS gels. Reference (2 papers): Steinberg et al., *Anal. Biochem.* **239**:223, 238 (1996).

There is also a procedure for specific staining on gels of glycoproteins, by the so-called periodate-Schiff reaction. Carbohydrates with OH on neighboring carbon atoms are oxidized to aldehydes, and the Schiff reagent, pararosaniline, reacts with these. The reference, regularly rediscovered, is a paper by Fairbank et al in *Biochemistry*, vol. 10 p. 2606, 1971.

Enzymes can often be located on gels by their activity, by generation of a colored or decolorated band in the gel, which means that their position can be identified and their mol. wt. determined even if they are not pure. Also, you may get several bands from 'pure' enzyme but find that they are all active. Activity stains are particularly important for observing isozymes in crude extracts, as in genetic studies.

There are several articles by Gabriel in *Methods in Enzymology* on enzyme staining. He uses several general terms for types of assay:

'Autochromic' methods, in which the enzyme product is colored on the substrate loses color - examples are production of *p*-nitrophenol from *p*-nitrophenyl glycosides, and decolorization of cytochrome *c* or hemoglobin included in the gel when it is prepared; a protease digests the protein and allows the heme to diffuse away.

'Simultaneous capture' in which the reaction product immediately reacts with other compounds present to generate color. The paramount example is the reaction of NADH or NADPH with phenazine methosulfate, which in turn reduces Nitro Blue tetrazolium to generate a purplish black band of insoluble formazan where the enzyme is. An insoluble band is obviously desirable.

'Post-incubation coupling' involves a second incubation under different conditions to generate the color. For instance, a keto sugar product such as TDP-4-keto-6-deoxyglucose, or just fructose, reacts with triphenyltetrazolium in 1 N NaOH at room temperature to give a pink color. Esterases and glycosidases can be located by using β -naphthyl esters/glycosides and diazo coupling the product β -naphthol with naphthylethylenediamine in nitrous acid. This is a much less soluble and more sensitive product than *p*-nitrophenol.

'Indicator gel' methods are used when substrates or coupling enzymes cannot be included in the running gel. The enzyme-containing running gel is sliced lengthwise - Scopes has a picture - and laid on an 'indicating' gel which contains the substrates and, if necessary, coupling enzymes. For instance, endo- β -glucosidases are located by laying the running gel on an indicator gel containing carboxymethylcellulose. After a period of incubation the running gel is removed, the indicator gel washed to remove hydrolyzed carboxymethyl cellulose, and then flooded with the indicator dye Congo Red, which reacts with intact carboxymethylcellulose to give a stable red color, but not where the carboxymethylcellulose has been digested, giving clear spots against a red background.

If you cannot think up a sensitive method for visualizing enzyme activity in the gel or on an indicating gel, you can always cut the gel into small slices, mash them to try to release the enzyme, and assay them separately.

Generally when you think you have a pure protein you stain a gel for protein, to show that you have only one band (you hope), and if possible for activity, to show that the one protein band seen *is* the enzyme. People have purified proteins until they couldn't see a band, but still had activity! But this was probably a deficiency in their protein stain. If possible you should electrophorese under two thoroughly different conditions, usually native and SDS gels, in case there is an impurity which happens not to be separated under one condition. Cationic native gels have also been used here, especially if the enzyme has an acidic pH optimum, but the rate of ammonium persulfate-catalyzed polymerization is pH-dependent, and at acidic pH photopolymerization with riboflavin has to be used, which is trickier.