Protein Purification

FINAL

William W. Ward* and Gavin Swiatek

Rutgers, The State University, School of Environmental and Biological Sciences, Department of Biochemistry and Microbiology, 76 Lipman Drive, New Brunswick, NJ 08901-8525, USA

Abstract: The science, art, and practice of protein purification have been with us for more than a century, yet, in many respects, the field is only now evolving past its adolescent roots. New methods are replacing old methods at such a dizzying pace, that even life-long experts in protein purification cannot keep up. In this article, we present many state-of-the-art protein purification techniques without totally ignoring the past. Our goal is to enable those relatively new to the field of protein purification to choose the best methods to solve their own purification problems. Each method we describe has been used and validated in our own research. We describe these methods, pointing out advantages, disadvantages, and limitations with practical examples rather than with complex, theoretical equations. This paper covers methods of extraction, clarification, batch purification, low pressure column chromatography, HPLC, and electrophoresis as applied to both genetically engineered, recombinant proteins and proteins isolated from natural sources. The relatively new methods of three-phase partitioning, hydrophobic charge induction chromatography, immobilized metal affinity chromatography, and perfusion chromatography are featured.

Keywords: Protein purification, extraction, clarification, chromatography, HPLC, electrophoresis, filtration.

INTRODUCTION

A Johns Hopkins professor, Maurice Bessman, who, many years ago, taught me protein chemistry, had a practice of saying, as an apparent quote from Efraim Racker, "Don't waste clean thinking on a dirty enzyme." He was, of course, talking about the need to purify a protein to homogeneity before spending time and energy studying its properties. A one percent contaminant in an otherwise highly purified preparation, for example, can give rise to incorrect or misleading observations and conclusions. The most extreme case of a "dirty enzyme" we have witnessed is one in which the "pure enzyme," described in great detail in a prestigious journal [1] was, in itself, a one percent contaminant of a 97% pure, "mystery" protein. That 97% pure "mystery" protein still has not been identified, even decades after the publication appeared.

Getting a protein pure is not easy! The process is as much an art as it is a science. As with music, architecture, and dance, protein purification follows basic rules, protocols, and practices. But, as is common with many artistic endeavors, protein purification relies heavily upon imagination, intuition, critical thinking, trial and error experimentation, and years of accumulated knowledge—both theoretical and practical. For these reasons, any introductory paper on protein purification must be viewed as a preliminary guide at best. In spite of these realities, we attempt, in this paper, to share with the readers some of our 65+ years of accumulated experiences working with proteins and developing ways to effect their purification.

*Address correspondence to this author at the Rutgers, The State University, School of Environmental and Biological Sciences, Department of Biochemistry and Microbiology, 76 Lipman Drive, New Brunswick, NJ 08901-8525, USA; Tel: 732-932-9763 x216; Fax: 732-932-3633;
E-mail: crebb@rci.rutgers.edu

PROTEIN CLASSES

Protein purification methodologies are currently applied to three distinctly different classes of proteins: (a) naturally occurring proteins, (b) recombinant proteins produced by introduction of appropriate genes into micro-organisms or cultured eukaryotic cells, and (c) complex mixtures of proteins arising from the application of genomics—a field, now called proteomics, that employs high throughput screening methods to simultaneously identify and characterize structures and functions of hundreds or thousands of gene products without necessarily purifying any one of the proteins to homogeneity. Rather than deal with the large scale screening questions addressed in the field of proteomics, we will concentrate, in this paper, on purification methods commonly used to purify naturally occurring and recombinantly derived proteins one at a time.

Technical advances over the past two decades, including optimization of gene expression in *E. coli,* Pichia, yeast, CHO cells, and other heterologous systems, and corresponding advances in downstream processing, have made protein production by genetic recombinant means the method of choice, when possible. As compared with methods for purifying proteins from natural sources, methods employing (suitably optimized) genetic engineering create more protein of interest with fewer contaminants. Recombinant tools provide elegant methods to achieve rapid, efficient purification via genetically introduced affinity tags. While one-step affinity purification methods are highly desirable, relatively few recombinant proteins can be purified to 99% purity by one-step methods. One or more additional steps are usually required. While affinity purification of recombinantly derived proteins has had a major impact, proteins that must undergo post-translational modifications (glycosylation, chromophore insertion, N-terminal modifications, etc.) to become func-

1573-4110/09 $55.00+.00

tional (an estimated 50% of all known proteins are post-translationally modified) are not as amenable to genetic cloning. In particular, heme-containing chromoproteins (like cytochromes, hemoglobin, myoglobin, and peroxidases) are difficult, if not impossible, to clone, as multiple gene products (enzymes) are required to assemble the heme tetrapyrrole ring. Recombinantly derived proteins that must be glycosylated to function have to be expressed in cells that add the correct oligosaccharides to the correct target amino acids. Bacteria fail to add any oligosaccharide units to target amino acids (specific serine, threonine, and asparagine residues) and yeast, while capable of glycoslyating proteins, generally do so in non-native ways. To effect precise glycoslylation patterns, recombinant proteins must be cloned into higher eukaryotic cells—often into cultured mammalian cells.

ASSAY METHODS

Before embarking on any protein purification project, it is necessary to select an appropriate biochemical assay to detect and quantify the protein of interest. A specific, identifiable property of that protein must be chosen as the assay basis. If the protein is an enzyme, the assay of choice measures that protein's ability to convert a substrate to a product, where, for signal-to-noise reasons, appearance of product is preferred over disappearance of substrate. When the protein of interest lacks enzymatic activity, that protein may be quantitated immunologically (typically by Western blot) or by densitometry (measurement of a stained band at the appropriate molecular weight position on an SDS polyacrylamide gel). With chromoproteins (eg. GFP, heme proteins, flavoproteins), protein quantity may be judged by measuring light absorbance at the wavelength maximum of the chromophore and purity may be judged by the ratio of chromophore absorbance to absorbance of total protein at 280nm [2]. The ratio (specific activity) of enzyme activity units (or chromophore absorption value) divided by a measure of total protein content is then used to estimate relative purity at all stages of purification. While it is assumed that purity will increase at each step of purification, there is no defined maximum value for specific activity. Rather, the specific activity, as a function of the number of purification steps, is expected to follow (to some extent) a hyperbolic curve, reaching an asymptotic value when the protein is pure. That the specific activity has reached a maximum must be verified by many independent ways. Accepted criteria include (a) constant specific activity across a chromatographic peak in the final, high resolution chromatogram "Fig. (**1**)", (ratio of curve 2 to curve 1 is represented as a dotted line) (b) a single band on an SDS-PAGE gel "Fig. (**2**)" (protein standards on the left, pure protein on the right), (c) a single symmetric peak on a size exclusion HPLC elution profile "Fig. (**3**)", (d) a single band on an isoelectric focusing (IEF) gel "Fig. (**4**)" (single bands in lanes 1,2,4,5,7, & 8 from left to right: standards in lanes 3 & 6), (e) N-terminal amino acid analysis showing just one amino acid, and (f) a single, unambiguous molecular weight determination by MALDI-TOF mass spectrometry (not shown).

Quantitating the denominator of the specific activity expression (total protein mass) is no easier than quantitating the numerator. Measuring total absorbance units at 280nm (opti-

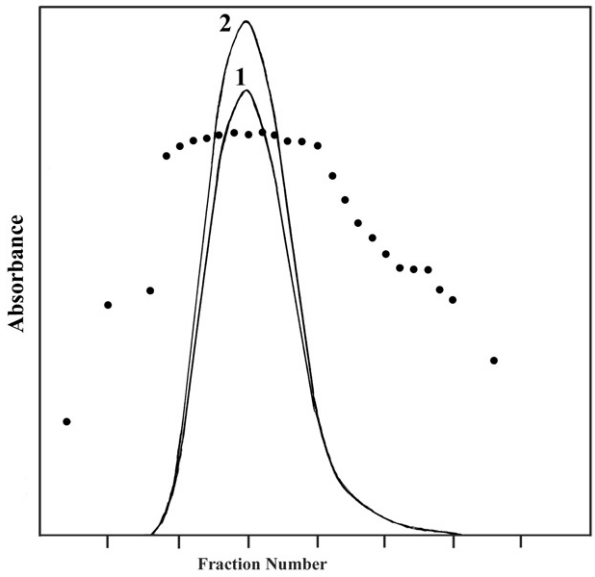

Fig. (1). Constant specific activity across a chromatographic peak. Here the absorbance at 397nm (2), a measure of GFP concentration, is divided by the absorbance at 280nm (1), a measure of total protein concentration, giving rise to a nearly flat dotted line across both peaks. Only at the extreme fringes of this preparative gel filtration chromatogram does the specific absorbance ratio drop significantly.

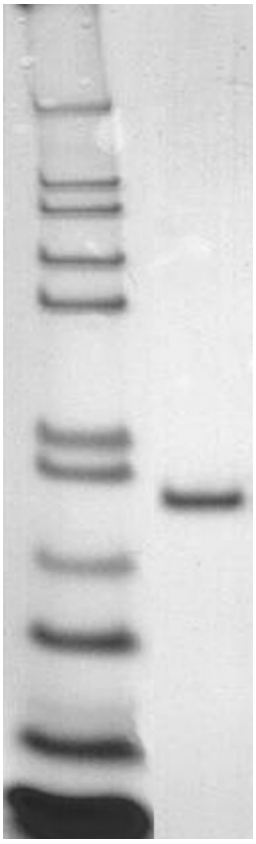

Fig. (2). Single band on an SDS-PAGE gel (right lane). Protein standards on the left lane. This criterion is often used as proof of purity, but it should always be accompanied by several other criteria. SDS gives approximate molecular weight of a monomeric protein or individual molecular weights of subunits of an oligometric protein.

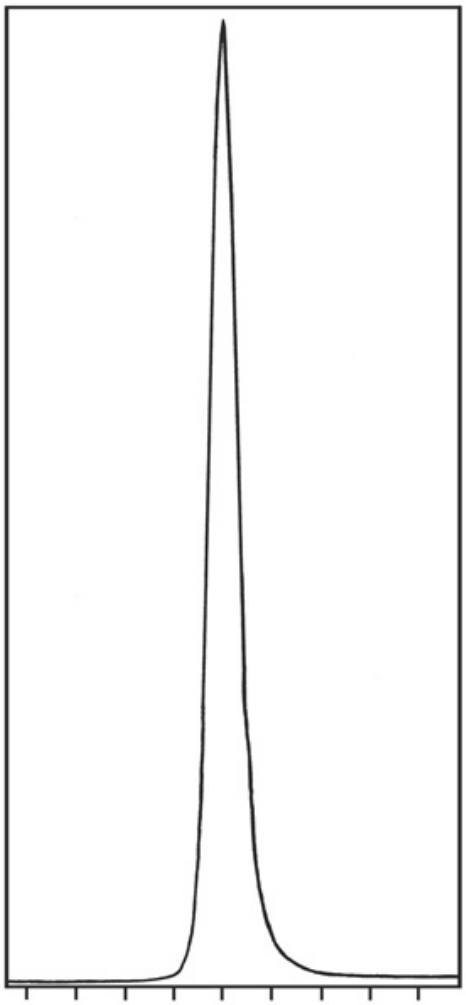

Fig. (3). Single symmetric peak on size exclusion HPLC. As this analysis is often performed on native proteins, it provides a molecular weight estimate for multisubunit proteins.

cal density X dilution factor X sample volume) is straightforward, quite reproducible, and very commonly practiced. However, that measurement assumes the same extinction coefficient (optical density/mg of protein/ml of solution in a 1 cm pathlength cell) for all proteins and no interfering absorbance from contaminents. However, the extinction coefficient of a protein at 280 nm depends almost entirely on the proportion of two aromatic amino acids—tryptophan and tyrosine—relative to all other protein components, including the other eighteen naturally-occurring amino acids and any oligosaccharides, lipids, and other post-translationally added groups. As these mass ratios vary as much as 20-fold from protein to protein and, as tryptophan is 5-times as absorptive as tyrosine "Fig. (**5**)", large errors in quantitating total protein mass by this method are very likely. Additional errors occur if samples contain, in addition to protein, DNA or RNA. These nucleic acids absorb maximally at 260nm but have strong spectral contributions at 280nm, as do polysaccharides containing uronic acids. In spite of all these uncertainties, a protein extinction coefficient of 1.0 absorbance unit/mg/ml is commonly assumed for protein mixtures.

Colorimetric assays for protein concentration may be utilized instead of direct UV measurements at 280 nm. The Lowry assay [3] is so commonly used that, for years, the original account was the most cited paper in all of science. One reason for the numerous citations is that countless substances have been reported in the literature [4] as interfering with the Lowry assay. Virtually every common buffer and buffer additive (phosphate, EDTA, mercaptoethanol, ammonium ion, Tris, etc.) interferes with the Lowry assay. The Bradford assay, marketed by Bio-Rad Laboratories, [5] is fairly sensitive, but this assay also has a long list of interfering substances and a wide range of colorimetric responses with respect to amino acid composition. An old stand-by is the Biuret assay [6], much less subject to interfering substances, but about an order of magnitude less sensitive than Lowry or Bradford assays. Probably the best colorimetric

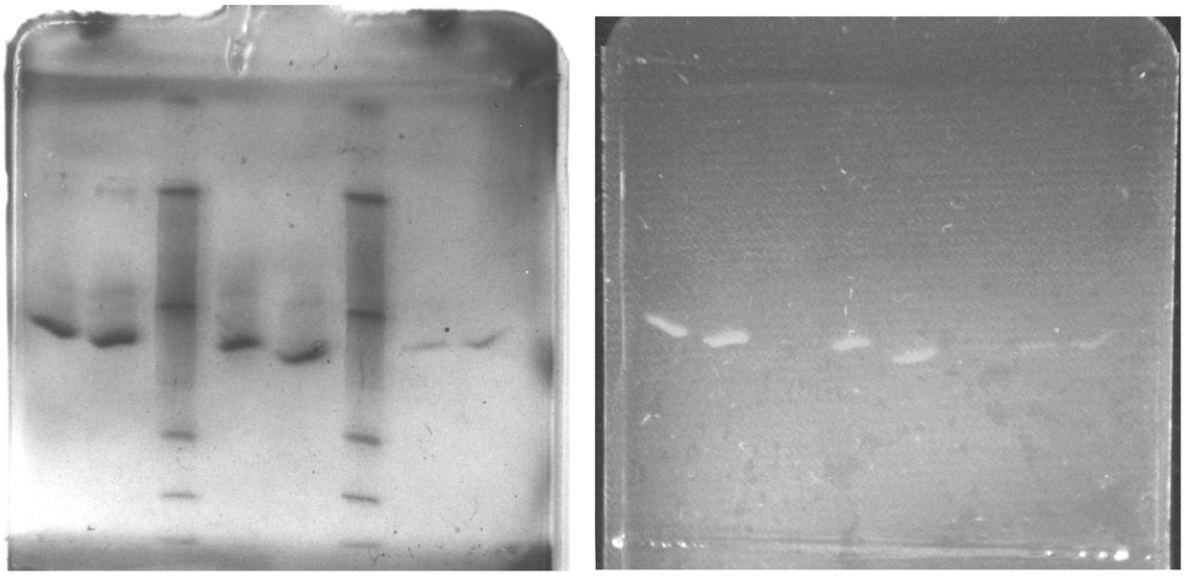

Fig. (4). Simultaneous IEF runs with GFP: Coomasie-stained gel on left, fluorescence imaged gel on right. Lanes 3 and 6 are protein standards

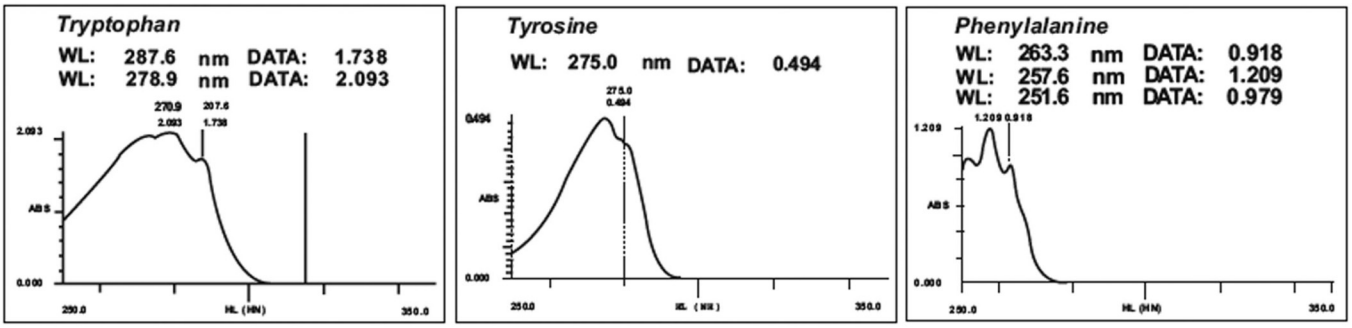

Fig. (5). Absorption spectra of the three aromatic amino acids found in proteins. Relative extinction coefficients at 280nm are 5 to 1 to 0 for tryptophan, tyrosine and phenylalanine, respectively.

assay, in terms of sensitivity and accuracy, is the BCA method sold by Pierce Chemical Company.

PROTEIN EXTRACTION

Proteins may be found: (a) free in solution (egg albumin, immunoglobulins, milk casein, secreted mucopolysaccharides), (b) sequestered within an extracellular matrix (seed coat proteins, oyster shell proteins, and proteins of the epidermal stratum corneum), (c) imbedded within phospholipid membranes (cell surface receptor proteins, visual pigment proteins, membrane transport proteins), (d) as components of connective tissue (keratin, yellow elastic fiber proteins, myosin, actin), or (e) contained within the fluid cytoplasm of cells (whether naturally occurring or recombinant). Some proteins associated with the extracellular matrix may be readily released by aqueous wash steps. Others require very vigorous extraction methods. Those imbedded within membranes usually require detergents or organic solvents for "solubilization," but they often remain as micellar emulsions rather than as truly soluble proteins. Intrinsically insoluble fibrous proteins that comprise certain connective tissues may need to be artificially solubilized by partial (or extensive) degradation with heat, acid, base, proteases, collagenases (matrix metalloproteinases) or strongly chaotropic agents like guanidine-HCl, urea, or detergents such as sodium dodecyl sulfate (SDS). Finally, water-soluble proteins within cells must be released by disruption of the cell membrane and, in the case of bacteria, fungi, and higher plants, by disruption of the cell wall as well as the membrane.

Methods for releasing water-soluble proteins from within cells vary with cell type. For many animal cells, hypotonic lysis [7] is sufficient to release soluble proteins. Hypotonic lysis is a two-step process—treatment of the cells with hypertonic salt to cause cytoplasmic shrinking (plasmolysis) followed by the plunging of those cells into distilled water. Rapid re-hydration of the cytoplasm causes the plasma membrane to burst, releasing the soluble cellular contents into the water. This otherwise very gentle hypotonic lysis method may, if necessary, be augmented by application of ultrasound bursts (sonication) or by homogenization with a Waring blender. In the latter two cases, DNA becomes sheared into protein-size pieces, lowering the viscosity but making DNA removal from protein more difficult.

Bacteria, fungi, and plants all contain cell walls that resist hypotonic lysis, making intracellular protein release more difficult. The proteoglycan cell wall of bacteria, especially gram positive bacteria, can be enzymatically cleaved by the action of egg white lysozyme [7]. Large quantities of this glycolytic enzyme (and long exposure times) are required, as lysozyme is strongly inhibited by cell wall cleavage products. Following cell wall disruption, release of intracellular proteins is accomplished by hypotonic lysis as above. If cell wall disruption is only partially accomplished, the process can be assisted by sonication or homogenization. Intracellular proteins of bacteria, fungi, and plants can be released with varying success by repeated cycles of freezing and thawing [7] or by vigorous sonication, each followed by homogenization as needed. The French press (10,000 psi followed by rapid return to atmospheric pressure) works well for all cell types, but is restricted to small volumes [7]. Larger volumes of cells can be disrupted by a bead mill [7], using the vigorous shaking of tiny glass beads to create highly disruptive local shear forces.

The more vigorous the method, however, the greater the quantity of unwanted cellular materials released. Furthermore, very large macromolecules (DNA and polysaccharides, for example) may be sheared into small pieces, complicating later size separation methods, as some of these fragments will now match the protein of interest in size and shape. Gentle lysis methods tend to release large, unsheared macromolecules and large lipid micelles that add to viscosity and sample turbidity, both of which interfere greatly with downstream processing.

The most effective cell disruption method we have used is a "one-step-does-all" method called three-phase partitioning (TPP) [8, 9]. In 30 minutes time, the three-phase method (see later description of TPP in the "Batch Purification" section) semi-selectively releases proteins from whole, unlysed *E. coli* cells while simultaneously removing all chromosomal DNA, all lipids, and the majority of unwanted proteins. TPP removes all particulates, lowers viscosity, and concentrates the resulting extract as much as 50-fold. While not suitable for all proteins (activity or solubility concerns are potential complications), when TPP works, it is unsurpassed by any other early-stage method.

BASIC STRATEGIES

The fundamental strategy of protein purification is very simple—to remove all contaminants while retaining as much as possible of the protein of interest.

1. Removing Other Macromolecules

Contaminants in protein extracts may include a variety of macromolecules (nucleic acids, lipid micelles, polysaccharides, and other proteins) as well as an assortment of small molecules. Small molecules are fairly easy to separate from proteins by size selection such as dialysis, ultrafiltration, and gel filtration. In the process of precipitating proteins with salts or other agents, small molecules usually stay in solution. Effective separation is then accomplished by centrifugation, causing precipitated macromolecules to collect as a pellet at the bottom of the centrifuge tube. Macromolecules are more difficult to remove, especially when present in large amounts. Lysed bacteria release large amounts of DNA, giving rise to extracts with high viscosity and high absorbance at 280 nm and 260 nm. DNA absorbs UV light at 260 nm about 20-40 times as strongly as protein absorbs UV light at 280, so the UV absorption spectrum of a bacterial extract often shows a peak near 260 nm "Fig. (**6**)" that overwhelms the protein absorbance at 280 nm. DNA can be removed by precipitation with protamine sulfate (a naturally occurring DNA binding protein containing 70% arginine and lysine) or by treatment with synthetic polyamine compounds. DNase and RNase may also be used to enzymatically depolymerize DNA and RNA, respectively. Coelenterates (jellyfish, sea pansies, and corals), from which green-fluorescent proteins are isolated, have large amounts of highly viscous proteoglycans (MW > 40 million). In one case, working with these slimy starting materials, it was discovered that 15 fractions, after eluting from an ion exchange column, had solidified like chilled agar. Fortunately, the solidification occurred after the chromatography run rather than on the column. When the latter happens, the column may be destroyed. For more than 20 years, we solved this viscosity problem by removal of proteoglycans on large (8 liter) gel filtration columns, that process consuming 4 person-months each year. Only recently did we discover that a simple pass through a very fast flowing column of diatomaceous earth (celite) accomplishes the same degree of viscosity reduction in less than 1% of the time. Egg yolk has very high levels of protein (~160 mg/ml) and high levels of lipids (neutral fats, steroids, and phospholipids). Only by applying unique methods, can the antibody, IgY, be isolated from egg yolk. When working with such difficult starting materials, attention must be given, early in the purification protocol, to removing problem contaminants, if only to lower viscosity, or to remove turbidity, or to improve optical clarity so that spectroscopic assays are not measuring mostly light scatter.

2. Purity vs. Recovery

Most preparative protein purification protocols involve several steps. Each has the potential to achieve a measure of purity, but narrow selection of the most highly purified fractions can greatly reduce yield. A protocol involving five purification steps, each one effecting a 5-fold purification, generates a potential overall purification factor of 3125-fold. But, if, in selecting only the best fractions, recovery of the protein of interest at each step is 70%, overall yield is just 16.8%. Increasing the step-by-step yield to 90% generates a final yield of 59%, but perhaps at the expense of purity. Thus, very significant overall loss of the protein of interest may come from selection of peak fractions at the expense of fringe fractions. Other losses come from time-dependent, irreversible binding to chromatography media, denaturation, protein oxidation, harsh binding or harsh eluting conditions from columns, heavy metal ion exposure, and separation of non-covalently bound cofactors, prosthetic groups, and stabilizing agents from the protein of interest. Sometimes a protein of interest exists in cells as an active non-covalent complex with one or more other proteins. During purification, the complex may be separated, leading to dramatic loss of activity (often reversible). A well-documented case is the separation of calmodulin from phosphodiesterase by ion exchange chromatography. As calmodulin facilitates the calcium ion stimulation of phosphodiesterase via close, non-covalent protein-protein interactions, separation of members of the interacting pair eliminates the calcium stimulation of phosphodiesterase [10].

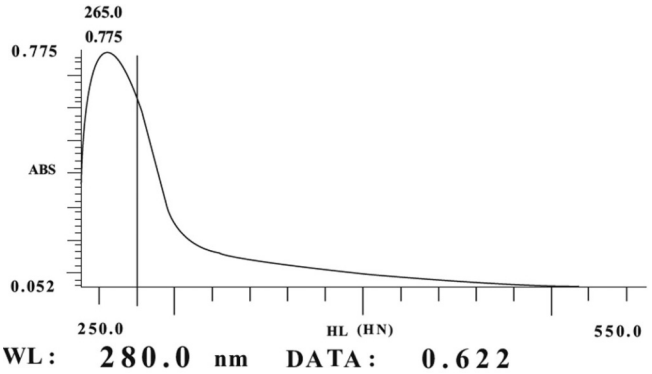

Fig. (6). The dominant absorbance in this crude extract of GFP-expressing *E. coli,* seen at 265nm, reflects a relatively high nucleic acid content. Subsequent purification of this same sample by TPP (Fig. 11) removes virtually all nucleic acid and so reveals the underlying protein absorbance, peaking near 280nm, and the now prominent spectral signature of GFP with peaks at 397nm and 475nm.

3. Protein Stability

Many proteins are highly unstable and some that appear very stable at one stage in purification may lose stability at another stage of purity, even under the same set of external conditions (pH, temperature, buffer composition, etc.). Thus, stability (activity in the case of enzymes) needs to be monitored at every stage of purification. A few general precautions are recommended for most proteins. Immediately after extraction, adding a variety of protease inhibitors (available from Roche Diagnostics as compatible mixtures of inhibitors in convenient pill form) will limit proteolytic clipping of the protein of interest. One of the key ingredients in many protease inhibitor cocktails is PMSF (phenyl methyl sulfonyl fluoride), a commonly used protease inhibitor that binds covalently to active site serine residues on serine proteases (trypsin, chymotrypsin, thrombin, subtilisin, etc.), permanently inactivating them. A metal chelator, ethylenediamine tetraacetic acid (1-10 mM EDTA) is used to bind heavy metals that, when free, can "poison" sensitive enzymes, activate certain metalloproteases, or enhance sulfhydryl oxidation. The reducing agents mercaptoethanol or dithiothreitol (DTT) may also be added at low concentration (1 mM) to prevent protein oxidation. Sodium azide (0.02%) prevents growth of

microorganisms, but azide cannot be used with peroxidases and other heme proteins without loss of enzymatic activity. Repeated sink-disposal of azide-containing buffers may cause the build up, on lead pipes, of explosive, and potentially deadly lead azide. Fisher Chemical sells Neutrazide, a chemical that will remove lead azide deposits on sink traps. It is a good idea to use Neutrazide, as plumbers don't like traps exploding in their faces! Ammonium sulfate (3-4 molar) is often used, during storage and between steps, as a powerful stabilizing agent for proteins. For this reason, protein supply companies often sell proteins as precipitated suspensions in saturated aqueous solutions of ammonium sulfate (4 molar). One should normally work at 0-4° C unless it is quite clear that the protein of interest is quite thermostable. On the flip side, many years ago, a few scientists working with large multisubunit proteins noticed that their proteins were much more stable at room temperature than in the cold. An explanation for this peculiar phenomenon had to wait a decade or more until hydrophobic interaction among and between proteins became more completely understood. The familiar thermodynamic equation ($\Delta F = \Delta G - T\Delta S$) offers the explanation. The multisubunit proteins in question are held together by hydrophobic bonds, the driving force for which is the entropy term "S." The net increase in disorder (entropy) comes from the release of structured water closely surrounding hydrophobic amino acid R-groups on protein subunits. When these subunits associate with each other, the structured water is released into the bulk solvent where it is more disordered. Increasing the temperature enhances the entropy term, favoring subunit-subunit interactions that stabilize the oligomeric protein.

4. Low vs. High Resolution

In general, the most efficient protocols begin with rapid, low-resolution steps that effect high recovery of the protein of interest and optical clarification of the sample (removal of turbidity). It is very helpful if the first step helps to stabilize and concentrate the protein of interest as well. Differential precipitation with ammonium sulfate followed by centrifugation often accomplishes each of these objectives [7]. To be effective, ammonium sulfate precipitation requires a minimum threshold concentration of total protein and protein of interest—1 mg/ml and 0.1 mg/ml, respectively, are reasonable guidelines. If the starting volume is very large, tangential flow ultrafiltration (TFF) can be used to reduce the volume prior to ammonium sulfate fractionation [11]. For dilute samples in small volumes, syringe-driven or centrifugal ultrafiltration units can be used in place of the large capacity TFF procedure.

BATCH PURIFICATION METHODS

1. Ammonium Sulfate Precipitation

Whether dealing with naturally occurring proteins or those genetically engineered, employing one or more early batch methods of purification is usually appropriate. For the purpose of isolating IgG antibodies, for example, treating whole rabbit serum with ammonium sulfate at 25% of saturation, followed by centrifugation, removes a small amount of proteinaceous precipitate. Further addition of solid ammonium sulfate, so as to achieve 37.5% of saturation, causes virtually all immunoglobulin (IgG) to precipitate while leaving nearly all of the albumin in solution after centrifugation. This very simple step achieves, at a minimum, a purification factor of 3-fold with almost no loss of IgG. Scopes [7] describes optimization of ammonium sulfate fractionation in great detail. Our experience is that such optimization is ineffective. With batch methods such as ammonium sulfate fractionation, it makes more sense to optimize yield rather than to optimize purity. It is best to use low resolution methods that yield high recovery of the protein of interest and then to employ high resolution methods to generate high purity in later steps.

2. Viscosity Reduction

Batch steps that remove DNA (and in doing so, lower the viscosity), like precipitation with protamine sulfate [7], are simple and straightforward. Other choices for lowering viscosity include adding nucleases or glycosidases to depolymerize DNA and viscous polysaccharides, respectively. For many years, in isolating GFP from natural sources, we reduced the viscosity of sea pansy extracts by adsorption onto alumina gels followed by centrifugation [12]. In the case of GFP isolation from *Aequorea* jellyfish, we used an 8-liter gel filtration column of P-100 BioGel. Using this column, more than 50% of the total A_{280} material (DNA and uronic acid-containing polysaccharides and proteoglycans) elutes in the void volume, while aequorin and GFP come out in later fractions, essentially free of all viscous molecules. Many years later, we discovered that passage of crude *Aequorea* extracts through Celite™ (diatomaceous earth) lowers viscosity just as effectively, doing so in minutes as opposed to two or three days with the P-100 column.

3. Acid Precipitation

A very simple batch method, used in very early stages of sea pansy GFP purification, is acid precipitation of unwanted proteins. Lowering pH of these extracts to pH 4.6 caused immediate, massive precipitation of contaminants. Performed at 4° C and followed by centrifugation at low temperature, the normally acid-labile GFP survives in solution with virtually no loss in endogenous fluorescence. Yet, 50% of the total contaminants precipitate and are removed during a brief high speed centrifugation step [2]. As previously stated, the conditions that define stability of any protein at one stage of purity may vary significantly at another stage of purity. Thus, *Renilla* GFP can be dropped to pH 4.6 in crude extract with no measurable loss in fluorescence, while the same protein, in the pure state, loses fluorescence rapidly as the pH is adjusted to 5.5 or below [2, 13].

4. Tangential Flow Ultrafiltration (TFF)

Tangential flow ultrafiltration "Fig. (**7B**)" is particularly well suited for concentrating and desalting large volumes of protein solution (liters to thousands of liters) [11]. Because fluid flow is rapidly directed across the surface of the membranes rather than perpendicular to their surfaces, TFF handles turbid feedstock as well as it does particle-free materials. In effect, tangential flow maintains a clean membrane surface by sweeping away particles that might otherwise clog the pores. If transmembrane pressures are optimized,

TFF can operate for many hours or days, at constant flow and constant pressure, recirculating the feed stock over the membrane surfaces. The volume of retentate slowly drops as low molecular weight solutes and water in the filtrate pass through the membrane and on to waste. Dead-end ultrafiltration, "Fig. (**7A**)", whether driven by hydrostatic pressure or centrifugal force is very convenient for concentrating and desalting smaller volumes.

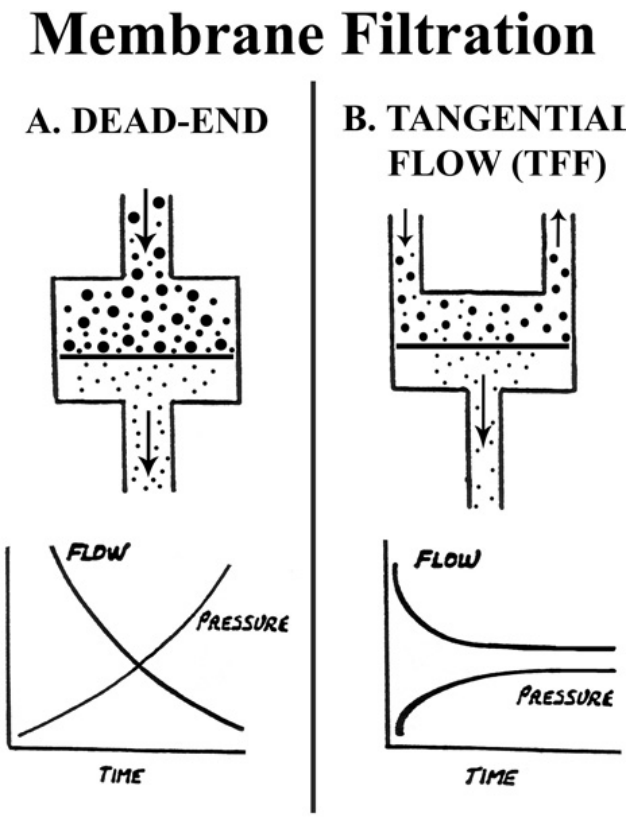

Fig. (7). Ultrafilters separate particles of molecular dimensions. The 10 kDa cut-off dead-end ultrafiltration unit pictured in **7.A**, traps most proteins while passing water, buffer salts and other small molecules. Flow rate drops and external pressure rises as proteins build up on the membrane surface. Tangential flow ultrafiltration, pictured in **7.B**, solves the membrane fouling problem by maintaining rapid flow across the membrane surface. Constant flow at constant pressure can be maintained for hours or, in some cases, days.

5. Three-Phase Partitioning (TPP)

Three-phase partitioning is a low-tech preparative (or analytical) method that accomplishes more than any other batch purification method we have utilized [8, 9]. In just a matter of minutes, TPP semi-selectively releases, purifies, and concentrates a protein of interest from whole, unlysed *E. coli* cells while removing all lipids, all chromosomal DNA, all pigments, nearly all viscosity, most polysaccharides, and most contaminating proteins. Requiring nothing more than a low speed centrifuge, TPP can isolate a desired protein from nearly all other protein and nucleic acid contaminants, as demonstrated by reduction, of up to 100-fold, in total absorbance at 280nm and 260nm, respectively. In the TPP process, viscosity and turbidity are virtually eliminated, producing an optically clear extract of the protein of interest at 50 to 100 times its original concentration.

Ideal for purifying GFP, TPP utilizes, in Stage I of the process, high concentrations of aqueous ammonium sulfate (0.8 to 2.4 molar, adjusted on a case-by-case basis for the protein of interest) together with an equal volume of tertiary butanol to effect selective precipitation of contaminants. Upon vigorous shaking to mix the two solvents, 10 to 20% of the water enters the organic layer, leaving the ammonium sulfate a bit more concentrated than it was in the first instance. Now, under the influence of the anti-chaotropic salt, t-butanol takes on a more hydrophobic character, self-associating to form a distinct organic layer that floats above the aqueous salt solution. The enhanced lipophilic character of tertiary butanol also makes t-butanol a much better solvent for membrane lipids. Virtually all the membrane lipids dissolve into the alcohol layer, opening the cytoplasmic contents of the cell to immediate access by the mixed solvent. Most cytoplasmic macromolecules, including chromosomal DNA and most proteins, precipitate *in situ*. Having exposed hydrophobic patches, these macromolecules bind large quantities of t-butanol, significantly lowering their buoyant densities. The large t-butanol-laden precipitates remain entrapped within the confines of the still-intact cell wall, unable to exit through the cell wall pores.

Low speed centrifugation produces three distinct layers: (a) an upper liquid layer of t-butanol containing a bit of water and nearly all lipids and pigment molecules, (b) a lower aqueous salt solution containing the protein of interest which, because the initial concentration of ammonium sulfate was pre-selected for that protein, has failed to precipitate, and (c) a pancake-like layer of precipitated nucleic acids, proteins, and cell wall polysaccharides—the densities of each having been greatly lowered by their having bound large amounts of t-butanol "Fig. (**8**)". To recover the still soluble protein of interest, the lower aqueous layer is retained while the upper and middle layers are discarded. Addition of an equal volume of fresh t-butanol followed by shaking (Stage II) further dehydrates, by another 5 to 15%, the aqueous layer, thus increasing the ammonium sulfate concentration proportionately. The elevated salt concentration forces the protein of interest now to precipitate, and, because that protein also binds a large amount of t-butanol, it is found floating above the aqueous salt solution as a very thin disk, following a few minutes of low speed centrifugation. In this photograph "Fig. (**9**)", the desired protein appears as a white layer sandwiched between the lower blue liquid (*E. coli* auto fluorescent material) and upper alcohol layer (appearing green-fluorescent via light-piping from below).

In Stage III of TPP, one decants away the two liquid layers, recovering the protein of interest as a thin disk that readily adheres to the side wall of the centrifuge tube "Fig. (**10**)". The disk is taken up in a minimum volume of aqueous ammonium sulfate identical in concentration to that used in Stage I and just sufficient to dissolve the protein of interest "Fig. (**10a**)". Centrifugation now reveals up to four phases in stage III - (a) a small volume of t-butanol floating on top, (b) a thin disk of precipitate just below, (c) a larger volume of aqueous ammonium sulfate containing the now soluble protein of interest, and (d) a small pellet of precipitated contaminant at the bottom of the tube.

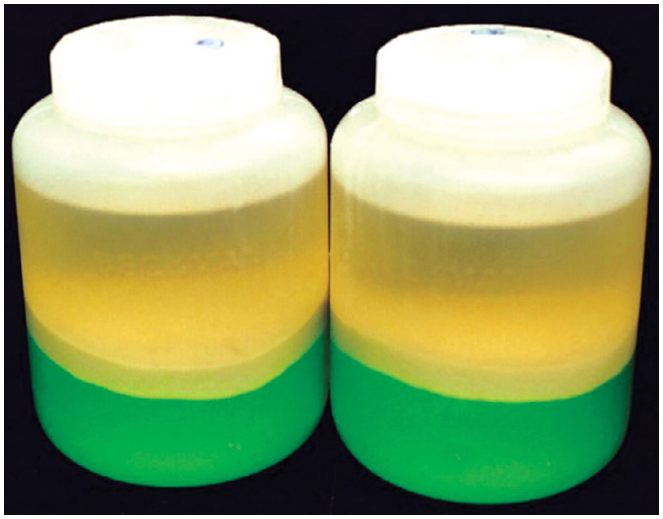

Fig. (8). Stage I of three-phase partitioning. The three layers are, from bottom to top, (a) GFP (green) dissolved in 1.6M buffered, aqueous ammonium sulfate solution, (b) nearly intact *E.coli* cells (cream-colored solid disk) containing most of the macromolecular materials, now precipitated and trapped within the undisturbed cell wall, and (c) t-butanol layer (diluted about 20% with water from the aequous layer) containing membrane lipids, pigments (yellow), and small hydrophobic molecules.

Fig. (9). Stage II of three-phase partitioning. The three layers are, from bottom to top, (a) further dehydrated aqueous ammonium sulfate solution containing blue-fluorescent materials from *E.coli,* (b) a 1-2 mm wide disk of now precipitated GFP, and (c) freshly added (to the green-fluorescent aqueous layer saved from stage I) t-butanol, appearing to fluoresce green along with the upper portion of the 750 ml centrifuge bottle via light piping from the fluorescence of the GFP disk.

Starting with whole, unlysed *E. coli* cells, the TPP procedure can be carried out in about 30 minutes, yielding up to a 100-fold purified protein of interest (with respect to absorbance at 280nm), concentrated at least 50-fold, and freed of viscosity, turbidity, total lipids, pigment molecules, and all chromosomal DNA. Fig. (**11**) shows the absorption spectrum of a TPP-purified lysate of *E. coli* that had been transformed with the gene for wild type recombinant GFP (note the characteristic GFP absorption peaks at 397nm and 475 nm). Compare this spectrum with that shown in Fig. (**6**) (before the administration of TPP). Not all proteins survive TPP. Some, like aequorin and soybean peroxidase, lose activity and some, like hemoglobin and myoglobin, fail to go back into solution in Stage III. Altering the protocol on a protein-by-protein basis (changing pH, organic solvent, or temperature, for example) may overcome some of these limitations.

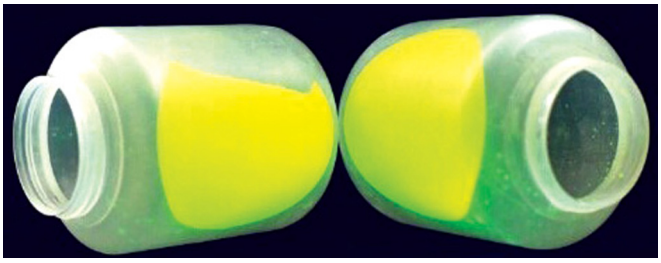

Fig. (10). Collected disk of GFP (appearing yellow with 365 nm illumination by virtue of self-absorbance of emitted fluorescence) from Stage II of TPP.

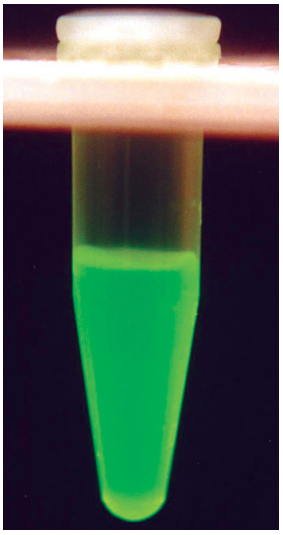

Fig. (10a). Stage III of TPP.

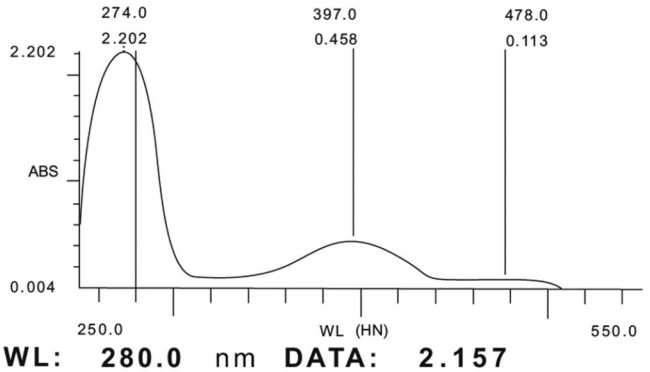

Fig. (11). Absorption spectrum of the now highly purified GFP sample following Stage III of TPP. Note the now prominent GFP "signature" peaks at 397nm and 475nm, unseen in Fig. 6 before TPP.

6. Miscellaneous Methods

Other batch methods [7] may include dialysis, ultrafiltration, and adsorption (at the proper pH and solvent condi-

tions) onto "loose" anion or cation exchange resins, hydroxylapatite or alumina gels, or hydrophobic interaction beads. Such methods can effect rapid, low resolution separation of the protein of interest from many contaminants that happen to have quite different chemical characteristics. Unlike axial flow column chromatography, none of these methods is greatly compromised by the presence of particles, micellar suspensions, or high viscosity. Once binding has occurred, sometimes gravity settling of the adsorption beads is all that is needed to remove significant levels of contaminants.

COLUMN CHROMATOGRAPHY

1. Introduction

With the exception of gel-filtration chromatography, all other column chromatographic methods are characterized as adsorption methods. Thus, under appropriate load conditions (buffer composition, pH, ionic strength, temperature, etc.) all adsorption methods have the ability to concentrate the protein of interest during the loading phase by factors as great as 100-fold or higher "Fig. (**12**)". To effect on-column concentration factors of this magnitude, the general practice is to establish loading conditions that favor very strong binding of the protein of interest. In the case shown in Fig. (**12**), a 60ml Phenyl Sepharose Fast Flow column in 1.6M ammonium sulfate has concentrated more than 2000ml of GFP during the loading phase of HIC chromatography. Under such conditions, many similar proteins will bind as well, competing with the protein of interest for binding space, the saturation limit often reaching 50 mg of total protein per 1 ml of gel medium [7]. Having bound many different proteins at the head of the column, the experimenter then selectively elutes the protein of interest by slowly changing the buffer elution strength. Usually this is done by establishing a shallow linear gradient (in the case of HIC, a gradient of decreasing ammonium sulfate).

2. Hydrophobic Interaction Chromatography (HIC)

For those familiar with C-18 reverse phase HPLC, the theory of hydrophobic interaction chromatography is quite similar. In both cases, the stationary phase is non-polar while the mobile phase is polar. Molecules with some hydrophobic character bind to the stationary phase beads during the loading phase by multiple van der Waals interactions "Fig. (**12**)". Release of binding is accomplished by decreasing the polarity of the mobile phases. In the case of C-18 RPC, the loading phase buffer is often 0.02% trifluoroacetic acid in water (pH 2) and the eluting solvent is an increasing gradient of methanol or acetonitrile. However, C-18 RPC is seldom used for proteins. The hydrophobic surface of reverse phase beads is usually very non-polar, with octadecyl chains bonded to silanol hydroxyl groups of silica gel in close proximity to one another. The densely packed octadecyl groups are able to "join forces" in binding to large proteins possessing multiple hydrophobic patches. These multiple interactions are so strong that proteins bind irreversibly. Manufacturers of C-18 RPC columns, appealing to the small molecule chromatography market, usually strive for very high "carbon loads" (high weight ratio of octadecyl groups with respect to SiO_2), making the binding of proteins to C-18 RPC beads even tighter. In addition, to generate high capacity binding for small

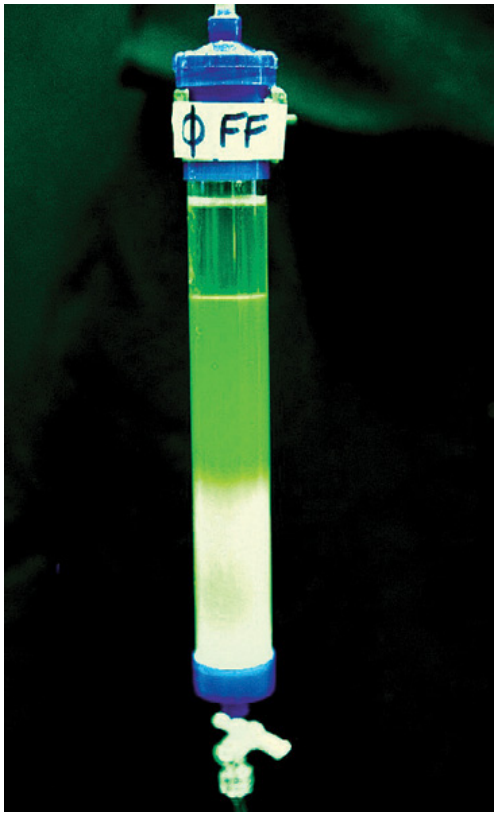

Fig. (12). Phenyl Sepharose Fast Flow chromatography of GFP. Here the loading stage of hydrophobic interaction chromatography (HIC) is shown. Approximately half of the HIC column (30ml) has been saturated with GFP, the column having trapped and concentrated GFP from a much larger volume (2000ml).

molecules, the manufacturers produce C-18 RPC beads with small pore sizes (80 angstroms is standard)—pores that cannot accept relatively large proteins into the interior of the beads. So, protein chromatographers tend to shy away from C-18 RPC, using instead high porosity agarose gels (300 angstrom pore sizes or larger) derivatized with butyl, octyl, or phenyl groups linked, in relatively low density, to the agarose sugar hydroxyls via 3-carbon-long spacers derived from the chemical linker epichlorohydrin, 6-carbon-long spacers derived from diaminohexane, or one of several other linkers. As the density of substitution with C-4, C-8, or phenyl groups is quite low, proteins become bound via fewer hydrophobic interactions than they would to a standard silica based C-18 RPC column. The relatively low intrinsic strength of binding permits controlled elution with aqueous solvents. In fact, the intrinsic strength of hydrophobic interaction between phenyl agarose (the most hydrophobic of the common HIC chemistries) and a typical water-soluble protein is sufficiently low that water, alone, as mobile phase, will not necessarily promote binding. So, to facilitate binding, it is necessary to use high concentrations of ammonium sulfate in the loading buffer "Fig. (**12**)". Release of the reversibly bound protein is accomplished by reducing the concentration of ammonium sulfate via continuous gradient elution or stepwise elution. The most hydrophobic of proteins may remain bound, even during the water wash. In such cases, a low concentration of organic solvent (10-20% isopropanol, for example) may be needed for final elution. Table **1** lists a

Table 1. Hydrophobic Interaction Chromatography Materials for Protein and Peptide Separation

Company	Product Name	Media Type	Ligand Density
GE Healthcare	Phenyl Sepharose™ 6 Fast Flow (low sub)	Phenyl Sepharose	25 µmol/ml medium
	Phenyl Sepharose™ 6 Fast Flow (high sub)	Phenyl Sepharose	40 µmol/ml medium
	Butyl Sepharose™ 4 Fast Flow	Butyl Sepharose	40 µmol/ml medium
	Butyl-S Sepharose™ 6 Fast Flow	Butyl Sepharose	10 µmol/ml medium
	Octyl Sepharose™ 4 Fast Flow[1]	Octyl Sepharose	5 µmol/ml medium
BioRad	Macro-Prep HIC Media Methyl HIC Support	Methacrylate-based 50 µm beads	70 - 125 µmol/mL
	Macro-Prep HIC Media t-Butyl HIC Support	Methacrylate-based 50 µm beads	70 - 125 µmol/mL
Tosoh	TSK-Gel Phenyl-5PW	Polymethacrylate	
	TSK-Gel Ether-5PW	Polymethacrylate	
	TSK-Gel Butyl NPR	Polymethacrylate	

[1]The Octyl Sepharose Fast Flow has a different hydrophobic character compared to the phenyl and butyl ligands, and is an important complement to the other hydrophobic matrices.

number of currently available HIC resins that illustrate a range of matrices, particle sizes and ligand chemistries.

HIC is a logical first column step for proteins that have undergone previous batch precipitation with ammonium sulfate. The starting sample, already in ammonium sulfate, may be safely stored in the cold for long periods of time in saturated ammonium sulfate. When the time comes to continue purification, the sample need only be thawed, adjusted to an appropriate ammonium sulfate concentration, and then loaded onto an HIC column equilibrated with the same concentration of ammonium sulfate—a concentration chosen to promote tight binding of the protein of interest. For many proteins, HIC provides significant chromatographic resolution while keeping the protein in a stabilizing medium. DNA and polysaccharides are likely to elute before most proteins. Lipids are likely to bind much more tightly than proteins. If early stage protein stability is a concern, HIC is an excellent choice for the first chromatography step, especially if fairly high ammonium sulfate concentration can be maintained throughout the process. If ammonium sulfate is to be kept at a high concentration, then a weakly hydrophobic medium (butyl agarose or methyl agarose) should be chosen. Using a weak HIC medium assures that the ammonium sulfate concentration can be kept very high, both for binding and for elution, so the stabilizing salt is always present. If early stage stability is not a concern, the strongly binding medium, phenyl agarose, may be selected. If, as often the case with phenyl agarose, it is necessary to elute with buffered water (or a low concentration of alcohol in aqueous buffer), the eluting sample may then be ready for immediate loading onto an ion exchange column, requiring, at most, a pH adjustment.

The history of hydrophobic interaction chromatography is particularly interesting. Instead of arising as the biochemist's version of reverse phase chromatography, HIC had an independent origin. In the 1970's after affinity chromatography was invented [14], some biochemists discovered that direct attachment of affinity ligands to agarose beads was ineffective in binding the enzyme they wished to purify. Suspecting steric hindrance, they re-designed the ligand attachment, using a spacer arm derived from diaminohexane. This 6-carbon-long bifunctional linker extends the "reach" of the bound ligand so that enzymes with deep active sites now have access to the ligand. While the spacer arm solved the access problem, it introduced a new problem. Binding specificity was not as good as it was designed to have been. Many protein contaminants lacking recognition sites for the bound ligand were binding to the affinity beads along with the protein of interest. This led to a control experiment in which the free end of an agarose-bound diaminohexane spacer arm was derivatized, not with the biospecific ligand, but with a methyl group. Switching from the biospecific ligand at the end of the diaminohexane to a simple methyl group was expected to create a chemically inert ligand. Instead, many of the same contaminants were found to bind to that methylated diaminohexane. This control experiment for affinity chromatography became the first HIC column.

3. Hydrophobic Charge Induction Chromatography

HCIC resins are an offshoot of thiophilic resins ("T-gels") that contain one or more free sulfhydryl groups capable of disulfide exchange with proteins having an exposed, reduced cysteine group. During development of the latter technology, in efforts to produce sorbents with differential specificity to immunoglobulins, aromatic terminal groups were added to existing "T-gel" supports. These terminal groups are most often nitrogen- or nitrogen- and sulfur-containing heterocycles. Resins containing terminal pyridines have been demonstrated to have the highest selectivity for immunoglobulins [15]. An example is MEP (4-Mercapto-Ethyl-Pyridine) HyperCel sold by Pall Corporation, in which

a strongly hydrophobic pyridinium group is immobilized to chromatography beads (M) with a sulfur-containing aliphatic linker arm. In its uncharged form (pH > 4), MEP resin selectively binds IgG. Lowering the pH to 4.0 (abruptly or via pH gradient) protonates the pyridinium ring, lowering its hydrophobicity and releasing bound proteins (Fig. **13**). A clear advantage of MEP over more conventional HIC matrices is that MEP binds proteins in the absence of added salt.

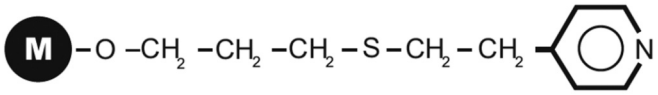

Fig. (13). Chemical structure of MEP ligand.

4. Ion Exchange Chromatography

Ion exchange is the oldest and most common protein chromatography method, having gotten its start with the Nobel Prize winning work of Moore and Stein in the 1940's [16] of separating amino acids on derivatized beads of polystyrene cross-linked with divinylbenzene. Two general forms of IEC exist, anion exchange (for negatively charged proteins) and cation exchange (for positively charged proteins). Most typically, protein mixtures are loaded onto ion exchange columns in low salt buffers at a pH that favors the protein of interest's having multiple negative charges (anion exchange) or multiple positive charges (cation exchange). Alternatively, the pH of the buffer can be adjusted so that certain classes of contaminants become bound to the column while the protein of interest washes through unretained. When the goal is to bind the protein of interest relatively tightly at the head of the column (so as to concentrate a dilute solution of the protein) the buffer concentration should be very low (1-10 mM) and the pH should be at least one unit below (for cation exchange) or one unit above (for anion exchange) the pI of the protein of interest. The pI is the isoelectric point of a protein—the pH at which the number of positive charges equals the number of negative charges.

Not including sulfonated and phosphorylated proteins (and other proteins having charged cofactors, metal ions, and prosthetic groups) the charges on proteins come from seven types of amino acid R-groups—those of glutamic acid, aspartic acid, histidine, lysine, arginine, tyrosine, and cysteine - and from the N-terminal amino group (pKa about 8) and the C-terminal carboxyl group (pKa about 2). Most proteins have an excess of acidic amino acid residues over basic amino acids, so these proteins are anionic (negatively charged) at neutral pH - their pI's are below 7. The higher the pH of the buffer, the more negatively charged a protein will be. However, with proteins in their native states, many potentially ionizable amino acid R-groups are not exposed to the external solvent. Their R-groups are buried in their uncharged states in solvent-inaccessible pockets within the protein interior. Many tyrosine residues in proteins are buried (in their protonated state and unresponsive to a rising pH until the protein denatures) [17].

Going from low pH to high pH, first the exposed glutamic and aspartic acids become deprotonated (negatively charged) as the pH rises above their pKa's of approximately 4.0). Next come the surface histidines, basic amino acids with a pKa near 7. Below pH 7, exposed histidine is positively charged. Above pH 7 it is neutral. Disulfide bonds are almost always buried in the interior of a normally folded protein, but even if on the surface, disulfide bonds cannot respond to pH unless first reduced by a chemical reducing agent. Free cysteines on the surface have a sulfhydryl group that ionizes from SH to S- above its pKa of 8. Much of the time cysteine is crosslinked to another cysteine by disulfide interaction. Lysine, the most exposed of all amino acid residues, loses its positive charge above pH 10, its pKa. The pKa for the phenolic hydroxyl of tyrosine is about 10, so above this pH, exposed tyrosines will carry a negative charge. Arginine loses its positive charge above pH 12, its pKa.

Modern ion exchange chromatography is capable of separating isoforms of the same protein that differ by a single charged residue on the surface of the protein "Fig. (**14**)". In the pictured case, 4 major isoforms of GFP have been resolved, the second one of which appears to be a poorly resolved doublet. Thus, with the exception of affinity chromatography, ion exchange offers greater potential resolution than any other chromatographic method. For this reason, we prefer to place ion exchange near the end of a purification protocol—after one or more batch steps and after HIC. If HIC precedes an ion exchange step and if the elution from HIC leaves the protein of interest in low salt, that protein may be ready for immediate application to an ion exchange column. At most, a pH adjustment may be necessary. For binding to anion exchange columns like DEAE (diethylaminoethyl) agarose, the exchanger and the sample must be equilibrated to a pH value at least one pH unit above the pI of the protein of interest (but never higher than 11, as DEAE loses its positive charge between pH 9 and pH 11). For proteins with very high pI values (or ones that, for stability purposes, must be kept at very high pH), quaternary ammonium ion exchangers may be utilized at pH values of 12 or higher.

For binding to a cation exchanger, the protein of interest must be titrated to at least one pH unit below its pI. The carboxymethyl group (immobilized acetic acid) is the cation exchanger most commonly chosen; however, CM groups (fully negative at pH values of 5 or higher) become protonated (and electrically neutral) below pH 4. For very acidic proteins, (pI values below 5), sulfonyl group exchangers or ones with immobilized phosphate groups are needed. Such cation exchangers, which remain charged at very low pH, are called strong cation exchangers. Exchangers that remain positively charged at very high pH are called strong anion exchangers. Whether anion or cation exchange is employed, selective elution of the protein of interest is most commonly accomplished by running an increasing, linear gradient of salt (NaCl) in the same equilibration buffer [7]. Positively charged buffers, like Tris-HCl, are utilized for anion exchangers while negatively charged buffers, like sodium phosphate or sodium citrate, are used with cation exchangers. Using positively charged buffer ions with negatively charged exchangers may cause large shifts in pH as the buffer, itself, binds to the exchanger. Recognize, as well, that negatively charged buffer additives (2-mercaptoethanol, sodium azide, and EDTA, for example) will bind to anion exchangers, building up to very high levels at low salt concentrations. Subsequent salt gradients will create substantial

Table 2. Ion Exchange Materials for Protein Separation

Company	Product Name	Ion Exchanger Type	Dynamic Capacity
	Q Sepharose FF	Quarternary ammonium strong anion	120 mg HSA/ml medium
GE	DEAE Sepharose FF	Diethyl aminoethyl weak anion	110 mg HSA/ml medium
Healthcare	ANX Sepharose 4 FF	Diethyl aminopropyl weak anion	5 mg thyroglobulin/ml med.
	SP Sepharose FF	Sulphopropyl strong cation	70 mg RNase/ml medium
	CM Sepharose FF	Carboxymethyl weak cation	50 mg RNase/ml medium
	Mustang Q[1]	Quarternary ammonium strong anion	85 mg BSA/ml medium
	Q Ceramic Hyper D	Quarternary ammonium strong anion	85 mg BSA/ml medium
Pall	DEAE Ceramic Hyper D	Diethyl aminoethyl weak anion	85 mg BSA/ml medium
	Mustang S[2]	Diethyl aminopropyl weak anion	75 mg lysozyme/ml med.
	CM Ceramic Hyper D	Carboxymethyl weak cation	60 mg IgG/ml medium
	Macro-Prep DEAE	Diethyl aminoethyl weak anion	35 mg BSA/ml
BioRad	Macro-Prep High Q	Quarternary ammonium strong anion	40 mg BSA/ml
	Macro-Prep High S	Sulphopropyl strong cation	70 mg IgG/ml
	Macro-Prep CM	Carboxymethyl weak cation	35 mg Hemoglobin/ml
	TSKgel BioAssist Q	Polymethyl acrylate strong anion	70 mg BSA/ml
	TSKgel QAE-2SW	Silica strong anion	ND
Tosoh	TSKgel DEAE-5PW	Hydroxylated methacrylate weak anion	30 mg BSA/ml
	TSKgel DEAE-NPR	Hydroxylated methacrylic polymer w. anion	5 mg BSA/ml
	TSKgel BioAssist S	Polymethylacrylate strong cation	70 mg BSA/ml
	TSKgel CM-5PW	Hydroxylated methacrylic polymer w. cation	45 mg Hb/ml

[1] Mustang Q membrane is a strong anion exchanger that effectively binds plasmid DNA, negatively-charged proteins, and viral particles; [2] Mustang S membrane is a strong cation exchanger that effectively binds positively charged proteins and viral particles.

peaks of these buffer additives in early eluting fractions, potentially interfering with commonly used protein assays [4].

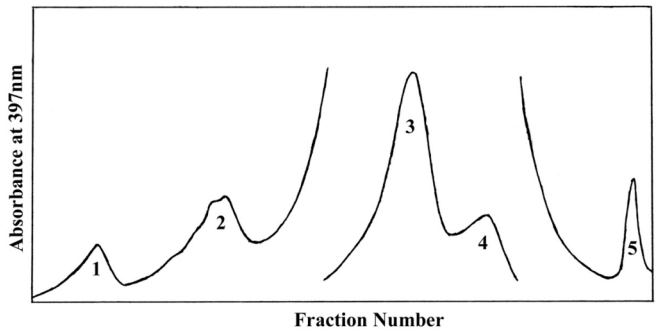

Fig. (14). GFP on DEAE Sepharose Fast Flow. Four GFP isoforms differing by just a few amino acids are clearly resolved in the shallow 4-liter gradient used to elute the sample. The peak labeled "5" results from a high salt wash that elutes "irreversibly" bound GFP.

5. Gel Filtration

While HIC separates proteins based upon the number and distribution of surface hydrophobic groups and ion exchange separates proteins based upon the number and distribution of surface charges, gel filtration (also known as size exclusion) separates proteins by size and shape. Gel filtration beads are similar to tiny sponges. Each gel filtration bead has numerous blind cul-de-sacs into which proteins diffuse—smaller, compact proteins penetrating farther than larger, irregular proteins. Compared with large proteins, small, compact proteins gain access to more pores and they travel farther into the channels. Thus, excluded from the many pores and channels, large proteins elute first [7].

Unlike adsorption chromatography, in which binding conditions cause dilute protein samples to concentrate on the bead surfaces at the top of the column, gel filtration is unable to concentrate protein samples. So, it is necessary for the experimenter to pre-concentrate samples. For maximum resolution in low pressure gel filtration, as demonstrated under optimal conditions in the Sephadex G-75 superfine column calibration profile depicted in (Fig. **15**), the column is long (~4 ft) and the sample volume is between 1 and 2 percent of the total column volume. The beads are very small (20-50 μM), flow rate is very slow (1-3 ml/cm^2/hr), viscosity

Table 3. Gel Filtration Materials for Protein Separation (Low to medium pressure chromatography)

Company	Product Name	Media Type	Working Range for Globular Molecules(MW)
GE Healthcare	Superdex Peptide PC 3.2/30	Agarose/Dextran	100 - 7000
	Superdex 75 PC 3.2/30	Agarose/Dextran	3000 - 70,000
	Superose 12 PC 3.2/30	Agarose	1000 - 300,000
	Superdex 200 PC 3.2/30	Agarose/Dextran	3000 - 600,000
	Superose 6 PC 3.2/30	Agarose	5000 - 5 x 10^6
	Sepharose 4 Fast Flow	4% Agarose	60,000 - 3 x 10^7
	Sepharose 6 Fast Flow	6% Agarose	10,000 - 4 x 10^6
	Sepharose 4B	4% Agarose	60,000 - 2 x 10^7
	Sepharose 6B	6% Agarose	10,000 - 4 x 10^6
	Sephacryl S-100 HR	Allyl dextran/bisacrylamide	1 x 10^3 - 1 x 10^5
	Sephacryl S-200 HR	Allyl dextran/bisacrylamide	5 x 10^3 - 2.5 x 10^5
	Sephacryl S-300 HR	Allyl dextran/bisacrylamide	1 x 10^4 - 1.5 x 10^6
	Sephacryl S-400 HR	Allyl dextran/bisacrylamide	2 x 10^4 - 8 x 10^6
	Sephacryl S-500 HR	Allyl dextran/bisacrylamide	4 x 10^4 - 2 x 10^7
	Ultrogel AcA 22	Polyacrylamide/Agarose	100 - 1200 KD
	Ultrogel AcA 34	Polyacrylamide/Agarose	20 - 350 KD
	Ultrogel AcA 44	Polyacrylamide/Agarose	10 - 130 KD
Pall	Ultrogel AcA 54	Polyacrylamide/Agarose	5 - 70 KD
	Ultrogel AcA 202	Polyacrylamide/Agarose	1 - 15 KD
	TrisAcryl GF05 M Grade	TrisAcrylamide	3000
	TrisAcryl GF05 LS Grade	TrisAcrylamide	250 - 2000
	Bio-Gel A Agarose Gel	Agarose	10,000 - 1.5 x 10^6
	Bio-Beads S-X Media	Styrene-Divinyl Benzene	400 - 14,000
	Bio-Gel P Polyacrylamide Media	Acrylamide/bisacrylamide	100 - 100,000
	Bio-Gel P 30 Gel Medium	N,N'-methylene-bis-acrylamide	2500 - 40,000
BioRad	Bio-Gel P 30 Gel Fine	N,N'-methylene-bis-acrylamide	2500 - 40,000
	Bio-Gel P 60 Gel Medium	N,N'-methylene-bis-acrylamide	3000 - 60,000
	Bio-Gel P 60 Gel Fine	N,N'-methylene-bis-acrylamide	3000 - 60,000
	Bio-Gel P 100 Gel Medium	N,N'-methylene-bis-acrylamide	5000 - 100,000
	Bio-Gel P 100 Gel Fine	N,N'-methylene-bis-acrylamide	5000 - 100,000
	G 2000 PW	Methacrylate resin	up to 2000
	G 2500 PW	Methacrylate resin	up to 3000
	G 4000 PW	Methacrylate resin	2000 - 3 x 10^5
Tosoh	SuperSW2000	Silica	5000 - 1.5 x 10^5
	SuperSW3000	Silica	10,000 - 5 x 10^5
	G2000SWxL	Silica	5000 - 1.5 x 10^5
	G3000SWxl	Silica	1000 - 3.5 x 10^4

is low, and fraction volume is small (no greater than 0.5 percent of column volume).

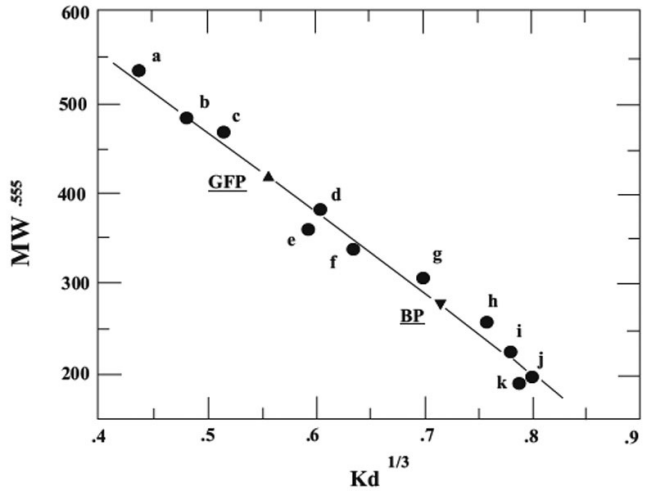

Fig. (15). Molecular weight determinations of Renilla GFP and Renilla luciferin binding protein obtained on a gel filtration column of superfine Sephadex G-75. Standard proteins used for these calibrations include: a, liver alcohol dehydrogenase; b, bovine serum albumin; c, hemoglobin; d, ovalbumin; e, horseradish peroxidase; f, β-lactoglobulin; g, carbonic anhydrase; h, trypsin inhibitor; i, myoglobin; j, ribonuclease; and k, cytochrome c. BP represents *Renilla* luciferin-binding protein, and GFP is the *Renilla* green-fluorescent protein. The plotting routine employed ($mw^{.555}$ vs. $kd^{1/3}$ helps to linearize an otherwise sigmoidal curve.

HPLC size exclusion columns achieve comparable resolution in 1% the time, but at a higher price and with smaller sample volumes "Fig. (**16**)".

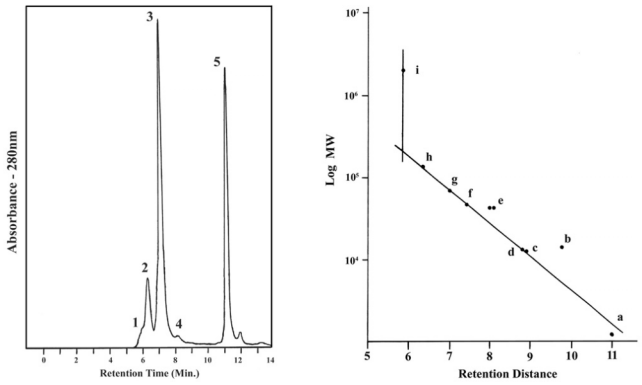

Fig. (16). Calibration of a Phenomenex HPLC BioSep SEC S-2000 column with protein standards of different molecular weights. Resolution is sufficiently great that bovine serum albumin, for example, can be resolved into three forms – tetramer (1), dimer (2 & h), and monomer (3 & g). Additional standards include: GFP (4), Vit B-12 (5 & c), lysozyme (b), CytC (c), RNase (d), SBP (e), Ovalbumin (f), & blue dextran (i).

6. Affinity Chromatography

Affinity chromatography is the most powerful chromatographic method known, as it exploits the biospecific interaction between protein and a covalently immobilized ligand [7]. The covalently bound ligand may be an enzyme inhibitor, a substrate analogue, a triazine-based fabric dye, an enzyme cofactor, an antigen recognized by a specific antibody, or a variety of other substances that form reversible complexes with the protein of interest. Elution conditions vary widely with the type of reversible complex formed. In the case of immunoaffinity chromatography, it is common to elute the reversibly bound substance (antibody or antigen) by lowering the pH to 3.0 in a strong citrate buffer (0.10M). Antibodies and many protein antigens denature at pH 3.0, disrupting the affinity-mediated binding and thus releasing the protein of interest from the column. Returning the pH to neutrality normally restores the antibody activity. When the interaction is absolutely specific (eg. antibody binding to a specific, unique, immobilized antigen) purification of that antibody can be accomplished in one step. However, because a bit of the affinity ligand may leach from the column, having formed a tight complex with the protein of interest, (in some cases as a multi-protein complex), it is advisable to pass the eluted protein of interest through a preparative gel filtration column to remove ligands that may have leached from the beads. Following preparative gel filtration, one can demonstrate removal of all high molecular weight multi-protein complexes by analyzing a portion of the final sample via size exclusion HPLC. A single, symmetric peak, as shown in (Fig. **3**), demonstrates complete removal of such aggregates.

7. IMAC (Immobilized Metal Affinity Chromatography)

The advent of recombinant DNA techniques in the 1980's made it possible to add affinity tags to cloned proteins of interest. Affinity tags in common usage include cellulose binding domain, maltose binding domain, glutathione binding domain, and polyhistidine tags (HIS-tags). Immobilized metal affinity chromatography (IMAC), introduced in 1975 by Porath, [18] employs a derivatized agarose bead to which metal chelating groups, such as iminodiacetic acid (IDA), or nitrilotriacetic acid (NTA), are bound. These groups can chelate nickel, cobalt, copper, or zinc ions in four locations on the metal, leaving two coordination spheres available for co-chelation by the protein of interest. Most naturally occurring proteins contain one or more surface histidines, cysteines, or tryptophans, enabling these proteins to co-chelate an immobilized metal ion. In doing so, these proteins will become reversibly bound to the beads.

This generic affinity method provides a means to partially resolve many naturally occurring proteins from each other, but without very high selectivity. Very high selectivity can be accomplished, however, by "tagging" the gene of interest with a string of 18 nucleotides coding for six sequential histidine residues. Such "His-tagged" proteins, as they are called, have so much greater affinity for the immobilized transition metal that they bind to the IMAC beads much more tightly than nearly all naturally occurring proteins "see Fig. (**17**)". Here GFP (green liquid) is being trapped on a Clontech Talon™ column of immobilized cobalt (red solid). Bound GFP appears yellow as the emitted light is self-absorbed by the highly concentrated column-bound GFP.

Often practiced by molecular biologists rather than by life-long protein biochemists, IMAC tends to be treated as a

simple, protocol-driven process. In following such protocols, experimenters bind His-tagged proteins to IMAC beads in cylindrical glass columns or onto loose beads that are collected in centrifugal "spin columns." Following a low-stringency wash step, the protein of interest is "bumped" from the column with a high concentration of imidazole, a compound that directly competes with the protein-bound His-tag for attachment to the immobilized metal.

Fig. (17). GFP on IMAC. Application of His-tagged GFP to a Clontech Talon™ column that utilizes cobalt ions to co-chelate the genetically added histidines on green-fluorescent protein. Illuminated with a 365nm UV lamp, unbound GFP appears green fluorescent, already bound GFP is yellow, and the bead-immobilized cobalt is pink.

Not uncommonly, a protein can be purified in this way to 95% homogeneity in a single IMAC step. But, by extending the wash step for 10 column volumes (or as much as 50 column volumes, in some cases) purity of 99+ percent can often be achieved. Adding a low concentration of imidazole (10-20 mM) to the wash buffer can shorten the wash step without compromising final purity. Following an extensive wash, the appropriate His-tagged protein of interest is now the last remaining protein on the column, as all other proteins in the mix contain insufficient surface histidines, cysteines, or tryptophans to effectively compete with a His-tagged recombinant protein. The His-tagged protein can then be "bumped" with imidazole-containing buffer at concentrations between 50 and 500 mM, the latter varying with the ligand density and the type of bound metal chelated by the agarose-NTA. Nickel, for example, requires higher imidazole concentration than cobalt.

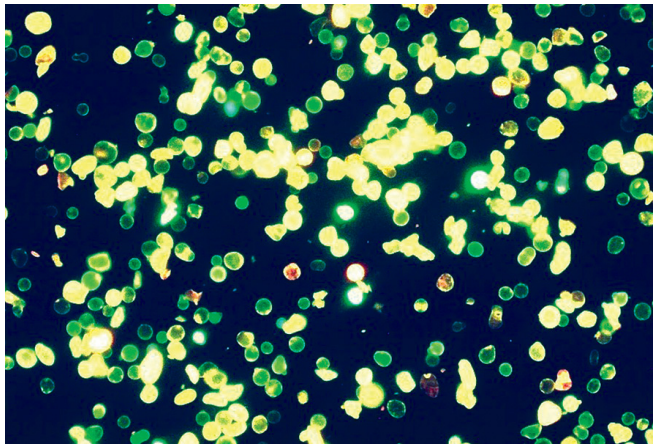

Fig. (18). GFP (green) and Ds-Red (yellow) fluorescent proteins bound to 5μM diameter C-18 reverse phase HPLC beads. Binding is so strong that GFP and Ds-Red are never seen to exchange places. These beads are very small (5 or 10μM) mono-disperse spherical particles of silicon dioxide, the silanol hydroxyl groups having been derivatized with 18-carbon-long (C-18) hydrophobic side chains to generate octadecylsilane (ODS) functional groups.

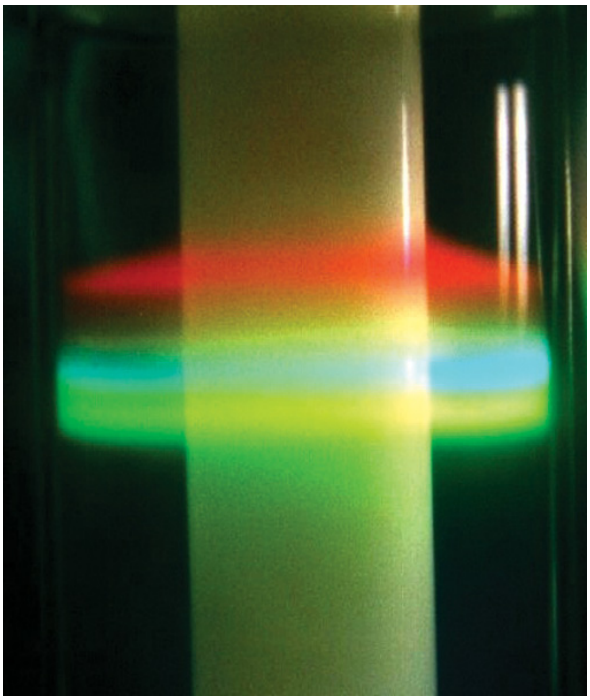

Fig. (20). Separation on the BioRad Prep Cell of coral-derived Ds-Red, Y66H (BFP), S65T (E-GFP), and several variants of wild-type rGFP with and without His tags. Electrophoresis is performed in an annular column of acrylamide under native conditions. A pH of 6.5 (close to the isoelectric points of the fluorescent proteins) maximizes potential separation of these bands. Eluted bands are collected in a fraction collector as they leave the column.

Important conditions must be met for successful IMAC purification. The buffer pH must always be at or above pH 7, as the imidazole group (on the His-tag and in the buffer)

must be deprotonated for effective chelation to occur. No free transition metals should be present in the buffers, as such metal ions may induce protein crosslinking through the histidines of two or more His-tagged proteins. The crosslinking problem becomes much more pronounced, perhaps leading to precipitation, when the recombinant proteins readily form dimers, tetramers, or other complexes. No other metal chelators (EDTA, EGTA, citrate, oxalate, NH_3) are permitted in the buffers, except in trace amounts. As it is a stronger chelator than NTA, EDTA, even in low concentrations, will readily strip the metal ions from the IMAC beads, preventing all binding of the His-tagged protein to the beads. In practice, we have used up to 2-3 mM EDTA in sample and wash buffers with little deleterious effect. IMAC may not be used directly following ammonium sulfate precipitation or hydrophobic interaction chromatography, as residual ammonia, in equilibrium with the ammonium ion, will prevent His-tag binding to IMAC beads.

In comparison with the highest quality gel filtration, ion exchange, and HIC media (that generally cost less than $5 per ml), affinity adsorbents are quite expensive (typically $20 to $100 per ml). For this reason, it is not advisable to introduce an affinity step before the feed stock has been clarified by high-speed centrifugation or ultrafiltration. Even following clarification, adsorption-induced protein concentration may cause on-column protein precipitation that can badly foul the column. In the worst case, an expensive affinity column may be ruined by such fouling. For this reason, we prefer to use affinity chromatography only after several other purification steps have been completed.

8. Covalent (Thiophilic) Chromatography

In nearly all cases of adsorption chromatography, attachment of a protein to the beads is via non-covalent chemistry. A notable exception is covalent chromatography. In this case, the sulfhydryl groups of one or more free cysteine residues on a protein become covalently bound, via disulfide linkage, to a bead-immobilized group also having a free SH group. That group is often thionitrobenzoic acid (TNB), from DTNB (dithionitrobenzoic acid). Covalent linkage of the carboxyl group of TNB to agarose hydroxyl groups is accomplished via coupling with a water-soluble carbodiimide. Following covalent binding, the attached protein can be eluted with a gradient of DTNB-containing buffer that participates in disulfide exchange with the bound protein, releasing the protein while recharging the column [19].

9. Beads and other Matrices

History

Early forms of chromatography utilized large (100-1000 μM), soft, porous, irregular beads, and, of necessity, operated at low pressure and low flow rate, delivering relatively poor resolution. With these older style chromatographic beads and fibers, most of the pores lead to blind cul-de-sacs so that analytes (whether chemically bound as in adsorption chromatography or passively retained as in gel filtration) remained "trapped" for relatively long periods of time in deep pockets and channels. These "trapped" proteins depend only upon static diffusion to provide their eventual access to the dynamic convective flow occurring outside the bead perimeter. In the case of adsorption chromatography, the eluting solvent would have to diffuse into the pores and channels, so as to weaken analyte binding to the internal adsorptive surfaces. Only then could the analytes diffuse to the outer surface and enter the convective flow. While channels of differing sizes and depths differentially "trap" analytes within the beads, external convective flow sweeps forward those analytes first to diffuse to the bead perimeter. Differential exit from pores and channels into the convective stream causes band broadening and, consequently, poor resolution.

HPLC

By reducing the diffusion distance between the adsorptive inner surfaces and the bead exterior, chromatographic performance (efficiency) in high performance liquid chromatography (HPLC), is markedly improved with respect to low pressure LC utilizing much larger beads. This shorter diffusion distance is generally achieved by greatly decreasing bead diameter. Thus, the beads in HPLC columns are typically mono-disperse, rigid, spherical particles 5μM or 10μM in diameter "Fig. (**18**)". The beads still have blind cul-de-sacs, but, now, diffusion distance from the bead interior to the external surface is reduced by an order of magnitude or more, producing much better sample resolution. But, the smaller beads restrict flow to such an extent, at near atmospheric pressure, that a high pressure pump (one capable of generating up to 7,000 lbs/in^2) is required for reasonable flow rates. This very high back pressure would crush softer beads composed of crosslinked polysaccharides or polyacrylamide. Thus, rigid beads of silica gel (SiO_2), that tolerate very high pressures, are normally the medium of choice for HPLC. Further requirements include stainless steel columns, high pressure fittings, and medium pressure flow cells (as the pressure drops substantially after liquid exits the column). As small beads act as excellent particle filters, HPLC columns are easily clogged unless buffers and samples are carefully filtered through fine pore particle filters (0.22 μM pores). Other disadvantages of HPLC include very high initial equipment investment (~ $100,000), relatively short column life, and relatively small sample capacity with standard size analytical columns (0.02 – 2.0mg).

An additional benefit of using high performance columns is that excellent separations can be realized in minutes of chromatography time as opposed to hours (or days) with low pressure columns. In addition, recovery of enzymatic activity is often greater with HPLC than with much slower low pressure columns. This is because, with some proteins, there is an inverse relationship between the recovery of activity and the length of time that a protein is bound to an adsorptive surface. Because HPLC offers higher resolution than corresponding low pressure methods, HPLC is a very useful analytical tool. Minor contaminants not resolved by preparative low pressure chromatography methods often show up by HPLC. In our lab, we routinely apply small protein samples to a Phenomenex BioSep S-2000 or S-3000 size exclusion HPLC column as a way to verify purity following preparative chromatography. Following just 20 minutes of running time, we obtain a trace of the HPLC size exclusion (SEC) profile, a representative sample of which is pictured in Fig.

(**3**). When we use the SEC column to verify purity of a heterotetrameric IgG antibody, we expect to find a single symmetric band corresponding to a molecular weight of 160 kdal. In parallel, we perform SDS gel electrophoresis, expecting to find a 50 kdal heavy chain and a 25 kdal light chain but no bands in the 160 kdal region.

Membrane Exchangers

More recent high resolution alternatives to HPLC include membrane exchangers (FMC, Pall, Sartorius) in which conventional membrane filters are derivatized with functional groups so as to promote ion exchange, hydrophobic interaction, or affinity binding. As most membrane exchange devices utilize simple syringes or peristaltic pumps to direct flow at 90 degrees to the membrane surface, fouling by particles suspended in the feed stock remains a problem unless the membrane exchanger is preceded by a removable particle filter. Sartorius and Pall have solved the fouling problem by introducing tangential flow membrane exchangers. As with tangential flow ultrafiltration, apparatus designs for tangential flow membrane exchangers pass the feed stock rapidly across the surface of the membranes in a continuously recirculating mode. The transmembrane pressure and transmembrane flow are kept low while the cross flow is brisk. Feed stock recirculation can be continued until all active sites within the membrane pores have been saturated with bound protein. Following extensive washing, pump velocity is greatly reduced, the cross flow outlet is clamped off, and elution buffer is gently forced through the membrane pores, releasing bound protein in a very small volume (pulse elution). If higher resolution is desired, a buffer gradient can be used to elute the protein of interest, selectively.

Perfusion Chromatography

About 20 years ago, Dr. Fred Rainier developed a unique chromatography concept known as perfusion chromatography. In this iteration of flow-through chromatography, Rainier fused together sub-micron silica particles in such a way as to create moderate size beads (50μm) having numerous flow-through channels. His invention became the basis for establishing PerSeptive Biosystems, a company later assimilated into Applied Biosystems. Rainier's invention was to agglutinate sub-micron size beads in such a way that uniform spherical aggregates (50 μM in diameter) are produced. The manufacturing process leaves numerous flow-through channels in each 50 μM bead, such that convective flow almost completely replaces static diffusion as the principle mass transport process bringing mobile phase into close contact with stationary phase. Not only is resolution markedly improved, but pumping can be done with low pressure peristaltic pumps or simple gravity feed. Prior to its assimilation into GE Healthcare, Pharmacia developed a similar matrix, designated "fast flow." All GE Healthcare chromatographic media (except for gel filtration media—that cannot function with flow-through beads) are available in "fast flow" versions. Other chromatography media suppliers have developed their own versions of "perfusion" or "fast flow" media. The resolving power of this type of chromatography medium is so great that it is possible to do high resolution chromatography without an HPLC system. We have been able, for example, to separate, with baseline resolution, gram quantities of GFP isoforms (one to three amino acid substitutions) on a 100 ml low pressure, gravity-fed column in an overnight run on DEAE Sepharose Fast Flow (see Fig. **14**).

CHROMATOGRAPHIC THEORY

Books have been written on the theory of column chromatography, especially HPLC. Of all the theories, equations, and concepts developed for LC, several stand out as especially helpful in practical applications. One important concept is the term "theoretical plate." The term "plate" is borrowed from the science of fractional distillation in which efficiency of distillation is related to the number of "plates," or surfaces, in the vertical column. As HPLC columns do not have physically identifiable "plates," the term, "theoretical plates," has been created. As with fractional distillation, the efficiency of chromatographic separation can be mathematically related to the number of theoretical plates. The resolving power of any column is related to the number of theoretical plates measured under isocratic (constant mobile phase) conditions. The plate count (**n**) is described by the equation:

$$n = 16 \bullet \left(\frac{t_R}{W_b} \right)^2$$

Where t_R is the retention time from sample injection to the center of a peak and W_b is the width of the peak at its base as measured by the baseline intersection of tangents to leading and trailing edges of the peak. Numerator and denominator are expressed in the same units, so **n** is dimensionless. The greater the number of theoretical plates, the greater is the column efficiency. The value for **n** is dependent upon two fixed parameters, column length and bead size, increasing linearly with column length and increasing inversely with bead size. Other factors within the user's control affect plate count. So, up to certain limits, **n** increases as the sample load, solvent viscosity, and flow rate each decrease. The plate count increases with increasing temperature and it decreases with the diameter of the analyte molecule. Our Phenomenex BioSep SEC S-2000 size exclusion column (see Fig. **3**) is rated at 70,000 plates per meter. As the column is about 1/3 of a meter long and as we use it for separating proteins (rather than the much smaller nucleotides that Phenomenex uses for calibration), we routinely measure 8,000 plates at flow rates of 0.5 ml/min. Plate count decreases measurably when the column is run at 1.0 ml/min, but shows no improvement when flow rate is lowered to 0.2 ml/min.

Van Deemter Equation

The distance between theoretical plates is called the plate height (***h***). The smaller the plate height the more efficient the column. The van Deemter equation shows the relationship between flow rate and plate height as follows:

$$h = A + \frac{B}{u} + C \bullet u$$

According to this equation, the plate height is the sum of three factors labeled the A-term, the B-term, and the C-term. The A-term (also called the eddy diffusion term) is the sum of all flow-independent factors that contribute to the reduc-

tion of column efficiency. Included are bead irregularities, column voids and packing irregularities, agglutination of beads, pre-column and post-column mixing, flow cell artifacts, etc. With the exception of poor plumbing, bad connections, and excessively large dead volume in the flow cell (all within the user's control), the rest of the A-term problems are not correctable for a given column.

The B-term, also called the static diffusion term, relates to the time-dependent loss of resolution (band broadening) caused by Brownian motion (simple diffusion of molecules) as well as convective mixing caused by differential heating. The faster the column flow rate, the shorter the time available for static diffusion to blur resolution, so the sharper the resolved bands. For small molecules in low viscosity solvents (tripeptides chromatographed in acetonitrile, for example) in a standard 6 mm diameter C-18 reverse phase HPLC with bead diameter of 5 μM, an acceptable B-term can be realized with flow rates of 1-2 ml/min. But, for large molecules (multi-subunit proteins like immunoglobulins with molecular weights of 160 kdal) chromatographed in aqueous buffer, the static diffusion term becomes almost negligible. Macromolecules such as IgG have intrinsically low diffusion rates and water (the preferred solvent for most proteins) is more viscous than acetonitrile. Thus, most protein chromatographers tend to ignore the B-term of the van Deemter equation, focusing virtually all their attention on the C-term.

The C-term, also known as the mass transport term, is, like the B-term, related to molecular diffusion. But, while the B-term focuses entirely upon time-dependent, non-convective diffusion of analytes, the C-term is also related to the flow rate of mobile phase across the beads. Traditional HPLC beads have the vast majority of their chemically interactive surfaces, not on the bead perimeter, but within the porous interior. External pores lead to intricate networks of branching channels generally ending in blind cul-de-sacs. Close to the bead perimeter, convective mixing occurs as the mobile phase passes by the closely packed beads in complex, non-laminar flow patterns. But, within the labyrinth of internal channels, convective flow goes to zero. All motion of the analyte within those channels becomes diffusion controlled. The larger the analyte and the more viscous the solvent, the slower this diffusion. An antibody molecule, for example, that happens to diffuse into a binding site in the deepest and least accessible cul-de-sac will, during the column elution phase, lag behind other identical antibody molecules that have bound closer to the surface of the bead. This differential release of bound molecules leads to band broadening and decreased chromatographic resolution. The faster the mobile phase is flowing, the greater this differential release will be, and thus, the greater the band broadening. HPLC has partially solved the C-term limitations by reducing bead diameter to the practical limit of 5 μM. But, with these small bead diameters comes higher resistance to flow necessitating pumps that generate higher pressures. Because the C-term dominates in protein chromatography, very slow flow rates generate the highest resolution.

Resolution Equation

Resolution (**Rs**) in isocratic adsorption chromatography is a measure of the degree to which two peaks are separated. Resolution is described by the equation:

$$\mathbf{R}_s = \frac{(V_{r2} - V_{r1})}{1/2(W_{b1} + W_{b2})}$$

Where V_{r2} = retention volume of peak #2, V_{r1} = retention volume of peak #1, W_{b1} = width at the base of peak #1, and W_{b2} = width at the base of peak #2. The industry-accepted resolution of 2.0 translates into a separation of two Gaussian peaks by a stretch of baseline as wide as one of those peaks "Fig. (**19**)".

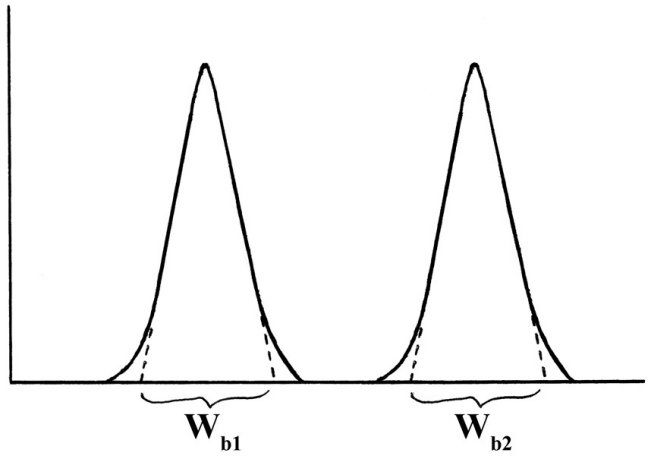

Fig. (19). Idealized chromatographic plot of two substances separated sufficiently to generate a resolution (**Rs**) of 2.0.

Resolution may be optimized by improving selectivity, sample retention, or column efficiency, the most important of these being selectivity. If the column is able to select between two analytes that differ greatly with respect to one characteristic (e.g. one analyte carries a net positive charge while the other is electrically neutral), then excellent separation can be achieved with a short runtime on a relatively inefficient column. Only when the two analytes are very similar in all respects must retention be increased and column efficiency optimized.

The equation describing optimization of resolution (**Rs**) is as follows:

$$\mathbf{R}_s = \frac{a - 1}{a + 1} \cdot \frac{\bar{k}}{1 + \bar{k}} \cdot \frac{\sqrt{n}}{2}$$

The three terms, from left to right, represent selectivity, retention, and efficiency. Where $\boldsymbol{\alpha}$ is the intrinsic chemical difference between two analytes, **k** is the degree of retention on the column, and $\boldsymbol{n}$ is the number of theoretical plates.

The most selective chromatography method applied to proteins is affinity chromatography. In its most discriminating mode, affinity chromatography can generate absolute selectivity. In such cases, where selectivity is very high, residence time on the column and column efficiency need not be optimized. But, when the ability of column chemistry to select between two or more very similar analytes (eg. the amino acids leucine and isoleucine) is very small, the other two parameters will need to be optimized.

ELECTROPHORESIS

Electrophoresis uses a relatively large voltage gradient (several hundred volts) to separate mixtures of proteins based on charge alone, charge/mass ratio, or size and shape. With some exceptions, notably proteomics (where the properties of electrophoretically separated proteins are analyzed in some detail), electrophoresis is used as an analytical tool rather than as a preparative tool, separating micrograms rather than milligrams of proteins. In such applications, the goal is to estimate protein purity or protein molecular weight. Moderate scale protein purification (1-50 mg, for example) by preparative electrophoresis, including isoelectric focusing, is possible, but it is never a simple matter. We've heard it said that every biochemist has at least one preparative electrophoresis unit somewhere in the lab on a high shelf collecting dust, because nobody in the group has been able to make it perform. The only unit with which we have had success is BioRad's Prep Cell, a preparative electrophoresis unit that separates multiple isoforms of a protein (single amino acid differences) in an acrylamide gel. The Prep Cell succeeds in maintaining excellent spatial resolution during the elution phase, delivering each isoform into a separate test tube. Capacity is about 1 mg per band. Fig. (**20**) shows the Prep Cell separating multiple forms of fluorescent proteins: DsRed, Y66H, S65T, and wild-type recombinant.

Analytical Electrophoresis

Six forms of electrophoresis dominate the field of analytical protein separations: SDS-PAGE, native PAGE, native acrylamide gradient PAGE, isoelectric focusing, two dimensional electrophoresis, and capillary zone electrophoresis (CZE).

SDS-PAGE

The most common analytical electrophoretic method, SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis), replaced analytical ultracentrifugation, almost overnight in the 1970's [20], as the "method of choice" in protein molecular weight determinations. Analytical ultracentrifugation is an expensive method generally requiring a dedicated technician in a core facility. SDS-PAGE, on the other hand, is a much less expensive technique that "everyone" can do. SDS (sodium dodecyl sulfate—also called laurel sulfate) is a powerful anionic detergent that binds tightly to proteins, denaturing them fully at elevated temperature (100°C) in the presence of a reducing agent like mercaptoethanol. Multi-subunit proteins (whose subunits are noncovalently associated or covalently attached by disulfide bonds) are dissociated into monomeric units by the combined action of SDS and a reducing agent at elevated temperature. So much SDS binds to a protein that the protein's intrinsic charge is overwhelmed by the bound SDS. In a crosslinked polyacrylamide gel (PAGE), electrophoretic mobility of a protein, so-coated with SDS, is almost entirely based upon size of the SDS-coated protein (a structure described as a rigid rod or as a detergent micelle). Monomeric molecular weight is then determined relative to standards of known molecular weights. Fixation of the slab gel with an aqueous methanol/acetic acid mixture, followed by staining with Coomassie Brilliant Blue R allows for the sample, (GFP IgG and IgY in various stages of purity) and a lane of standards (lane 6) to be compared—migration distance being inversely proportional to the log of the molecular weight "Fig. (**21**)". While almost always used in an analytical capacity, SDS-PAGE provides sufficient protein per band (1-10 ug, up to 200 pmoles) that protein from the sliced-out band of interest can be transferred to a PVDF membrane and processed directly for N-terminal protein sequencing via Edman degradation.

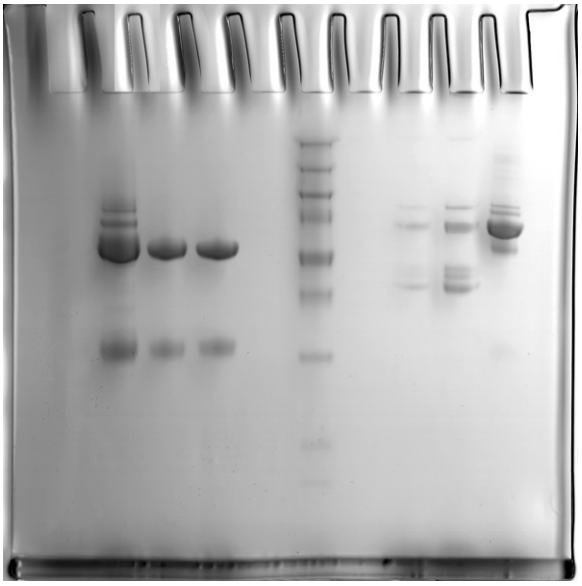

Fig. (21). Multi-lane SDS polyacrylamide slab gel showing lanes that had been loaded with partially purified anti-GFP IgG from rabbits (lanes 2-4), protein standards (lane 6), and partially purified anti-GFP IgY from chicken egg yolk (lanes 8-10). Lanes 5 and 7 were left blank to avoid "cross-talk." Coomassie dye was used to stain the gel after fixation in methanol/acetic acid.

Native PAGE and Acrylamide Gradient PAGE

Native gel electrophoresis separates non-denatured proteins in aqueous buffer, based on differences in the net charge of each protein, the chosen concentration of acrylamide being too low to effect size selection except for very large multi-subunit proteins. However, when the gel is prepared by polymerizing a continuous concentration gradient of monomeric acrylamide mixed with the cross-linking agent, bis-acrylamide (low concentration of acrylamide at the top of the gel and high concentration at the bottom) a gradient of pore sizes is produced upon polymerization of the gel. At alkaline pH, the voltage gradient accelerates all negatively charged proteins toward the anode, initially at rates dependent upon the protein's net charge. The large pores at the top of the gel fail to separate proteins by size, so mobility depends upon net charge alone. Soon, however, the largest among the proteins experiences diminished mobility as fewer pores permit passage. Eventually, these proteins cease migrating altogether. Smaller proteins, even those with lower intrinsic charge/mass ratio, overtake the larger proteins until they, too, come to rest in an acrylamide zone having pores too small for their further passage. Eventually all proteins come to rest as each encounters a prohibitive layer of pores. So the final migration distance is based entirely upon protein size and shape, the distance being a function of the log of

molecular weight for spherical proteins. As these gels are run under non-denaturing conditions, oligomeric proteins migrate without dissociation into subunits. Staining can be via Coomassie Brilliant Blue R, as with SDS gels. Or, because the proteins are expected to retain biological properties, they can be detected by immunological means or by enzymatic activity in which a substrate is converted to a detectable, colored product (a process that produces what is called a zymogram).

Isoelectric Focusing

Isoelectric focusing separates proteins exclusively on the basis of net charge, discriminating among proteins that differ in isoelectric point (pH at which the protein carries a zero net charge) by as little as 0.01 units [7]. To accomplish this fine discrimination, an acrylamide gel is polymerized with a complex mixture of polyamino-polycarboxylic acids knows as ampholytes. The ampholytes, being small in size and multiply charged, migrate quickly in the voltage gradient, sorting themselves out by their own pI values to create as many as 2000 individual pH zones within the gel. Strong acid and strong base at the anode and cathode, respectively, prevent the ampholytes from entering the electrode chambers. Once equilibrium is reached (as measured by current having reached a minimum) a stable pH gradient will have been established by the electrophoretic positioning of the ampholytes. Proteins migrate within this pH gradient until they, too, stop all net migration at their respective pI (isoelectric) values. A protein with a pI of 4.92 might come to rest between a pair of ampholytes having pIs of 4.89 and 4.93, effectively separated from another protein with pI of 4.94 and from another with pI 4.88. Such separations may be achieved on an analytical gel with post-electrophoresis fixation to remove the ampholytes and precipitate the proteins before application of the stain.

Two Dimensional Gel Electrophoresis

2D gel electrophoresis is frequently used in proteomics, where there is a need to separate and measure a very complex mixture of proteins. This method combines two individual methods, IEF and SDS-PAGE, to provide separation based on charge in one dimension and size in the other. Isoelectric focusing is commpnly performed on immobilized pH gradient (IPG) strips in the presence of the non-ionic denaturing agent, urea. After the proteins are focused along the length of the IPG strip, the strip is inserted perpendicularly in the top of an SDS slab gel and the interface sealed with agarose. When the voltage gradient is applied across the slab, each anionic protein in the tube is drawn into the SDS slab gel where SDS replaces the urea as the denaturing agent. Now the proteins, previously separated by charge via IEF on the IPG strip, migrate through the SDS slab at 90 degrees to the direction they had migrated during the IEF phase. The electrophoretic migration proceeds as in a normal SDS gel with normal fixing and Coomassie staining. The stained gel, however, has an exceedingly complex two-dimensional pattern of spots, allowing the experimenter to identify hundreds of unique proteins.

Capillary Electrophoresis

Almost always, capillary electrophoresis is an analytical method, as sample size is in the nanoliter range. While somewhat temperamental, CZE offers extraordinarily high resolution, especially in the IEF mode. Detection of focused protein bands is complicated, however, by the background absorption of the ampholytes. To overcome ampholyte absorption at 280nm, very high protein concentration is required.

PROTEIN MICROHETEROGENEITY

Native proteins with identical or near-identical properties, that are isolated from eukaryotic organisms, frequently display one or more amino acid substitutions, deletions, or additions. This microheterogeneity may be traced to the fact that higher organisms are diploid—they have two sets of chromosomes. Each one of a pair of homologous chromosomes has potentially different corresponding genes that give rise to different allelic traits in the adult. Higher plants may have many sets of chromosomes, increasing the chances of microheterogeneity in their corresponding gene products. Genetic drift within a large population of organisms may add further to the microheterogeneity. Thus, from a population of 100,000 *Aequorea victoria* jellyfish, one can identify up to 12 isoforms of the bioluminescent photoprotein, aequorin [21, 22], and up to 8 isoforms of the bioluminescence accessory green-fluorescent protein "Fig. (**22**)", all variants of which are assumed to have originated from very slightly different gene sequences in the population.

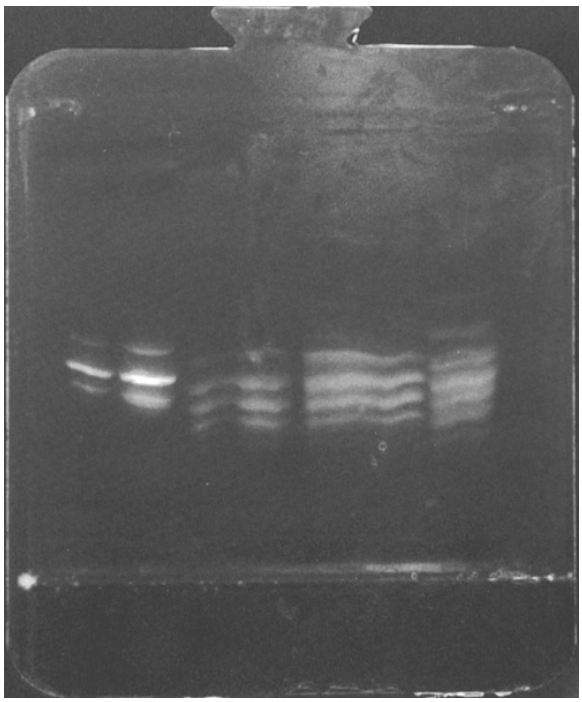

Fig. (22). Isoelectric focusing FAST™ gel (GE Healthcare) showing multiple isoforms of GFP imaged by fluorescence.

Additional microheterogeneity may occur during post-translational processing within a host cell (differential glycosylation for example) or may result from artifactual "damage" during extraction, purification, and storage of the protein. Common sources of post-translational microheterogeneity include deamidation of glutamine and/or asparagine, oxidation of cysteine, tryptophan, or methionine, and N-

terminal or C-terminal proteolytic clips from the action of endogenous peptidases. Differential glycosylation of serine, threonine, and asparagine residues is fairly common. The N-terminal amino acid may be differentially acylated while urea-exposed protein may experience differential carbamylation of lysine residues, reversing the charges on some of the epsilon amino groups and, in doing so, producing a complex array of isoforms.

Even in preparing recombinant proteins in haploid microorganisms like bacteria, there are non-genetic sources of microheterogeneity. Endogenous proteases in host cells may catalyze the proteolytic clipping of one or more histidine residues in a His-6 affinity tag. Binding to and elution from an immobilized metal ion column may fail to resolve recombinant proteins lacking one or more of the genetically programmed histidine residues. Most of these post-translational modifications create heterogeneity with respect to charge and molecular weight, complicating most analyses, including mass spec analysis.

CONCLUSION

In this article we have described a variety of protein purification techniques we practice in our everyday lives. To illustrate methods we frequently use, we have included graphs, figures and other forms of data that originated from our work. Where possible, we have explained the methods that work well for us and why we prefer these methods over others. We have tried to focus on practical methods and procedures rather than abstract theory, assuming that mastery of protein mass spectrometry is sufficiently demanding and time consuming that many practitioners of mass spectrometry may not also be experts in protein sample preparations. Those who wish hands-on experience in protein purification are invited to visit our website: http://www.rci.rutgers.edu/~crebb/protein.html.

ACKNOWLEDGEMENTS

"The authors wish to acknowledge the excellent work of Randy Ward who formatted the manuscript, created the charts, graphs and formulas, applied Photoshop techniques to edit and remove blemishes from the figures, and kept things going with his constant attention to detail."

REFERENCES

[1] Karkhanis, Y.D.; Cormier, M.J. Isolation and properties of *Renilla reniformis* luciferase, a low molecular weight energy conversion enzyme. *Biochemistry*, **1971**, *10,* 317-26.

[2] Ward, W.W.; Cormier, M.J. An Energy Transfer Protein in Coelenterate Bioluminescence: Characterization of the *Renilla* Green-Fluorescent Protein. *J. Biol. Chem.*, **1979**, *254*, 781-88.

[3] Lowry, O.H.; Rosebrough, N.J.; Farr, A.L.; Randall, R.J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem*., **1951**, *193*, 265-75.

[4] Ward, W.W.; Fastiggi. Binding of EDTA to DEAE-cellulose and its interference with protein determinations. *Anal. Biochem*., **1972**, *50*, 154-62.

[5] Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*., **1976**, *72*, 248-54.

[6] Gornall, A.C.; Bardawill, C.J.; David, M.M. Determination of Serum Proteins by Means of the Biuret Reaction. *J. Biol. Chem*., **1949**, *177*, 751-66.

[7] Scopes, R.K. *Protein Purification, Principles and Practice*; *3rd ed*.; Springer Advanced Texts in Chemistry, Springer-Verlag, N.Y., **1994**.

[8] Dennison, C.; Lovrien, R. Three phase partitioning: concentration and purification of proteins. *Protein Expr. Purif*., **1997**, *11*, 149-61.

[9] Thomson, C.M.; Ward, W.W. Three-phase partitioning (TPP): A rapid and preparative purification tool for GFP. In *Bioluminescence and Chemiluminescence*; Stanley, P.E.; Kricka, L.J. Eds.; World Scientific: N.J., **2002**; pp 115-18.

[10] Cheung, W.Y. Cyclic 3',5'-Nucleotide Phosphodiesterase - Evidence For and Properties of a Protein Activator. *J. Biol. Chem.,* **1971**, *246*, 2859-69.

[11] Strathman, H. Membranes and membrane processes in biotechnology. *Trends Biotechnol*., **1985**, *3*, 112-18.

[12] Charbonneau, H.; Cormier, M.J. Ca^{2+}-induced bioluminescence in Renilla reniformis. Purification and characterization of a calcium-triggered luciferin-binding protein. *J. Biol. Chem*., **1979**, *254*, 769-80.

[13] Bokman, S.H.; Ward, W.W. Renaturation of *Aequorea* green fluorescent protein. *Biochem. Biophys. Res. Commun*., **1981**, *101*, 1372-80.

[14] Cuatrecasas, P.; Wilchek, M. Specific enzyme purification by affinity chromatography. *Proc. Natl. Acad. Sci*., **1968**, *61*, 636-43.

[15] Oscarsson, B.; Porath, J. Protein chromatography with pyridine- and alkylthioether-based agarose adsorbents. *J. Chromatogr.*, **1990**, *499*, 235-47.

[16] Moore, S.; Stein, W.H. Column chromatography of peptides and proteins. *Adv. Protein Chem.*, **1956**, *11*, 191-236.

[17] Glazer, A.N. In *The Proteins,* Neurath, H.; Hill, B.L.; Boeder, C.L., Eds.; 3rd ed.; Academic Press Inc: New York, **1976**; Vol. 2; pp 1-103.

[18] Porath, J.; Carlsson, I. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, **1975**, *258*, 598-99.

[19] Matthews, J.C.; Hori, K.; Cormier M.J. Purification and properties of Renilla reniformis luciferase. *Biochemistry*, **1977**, *16*, 85-91.

[20] Weber, K.; Osborn, M. Reliability of molecular weight determinations by dodecyl sulphate-polyacrylamide gel electrophoresis. *J. Biol. Chem.*, **1969**, *244*, 4406-12.

[21] Blinks, J.R.; Harrer G.C. Multiple forms of the calcium-sensitive bioluminescent protein aequorin. *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, **1975**, *34*, 474.

[22] Shimomura, O. Isolation and properties of various molecular forms of aequorin. *Biochem. J.*, **1986**, *234*, 271-77.