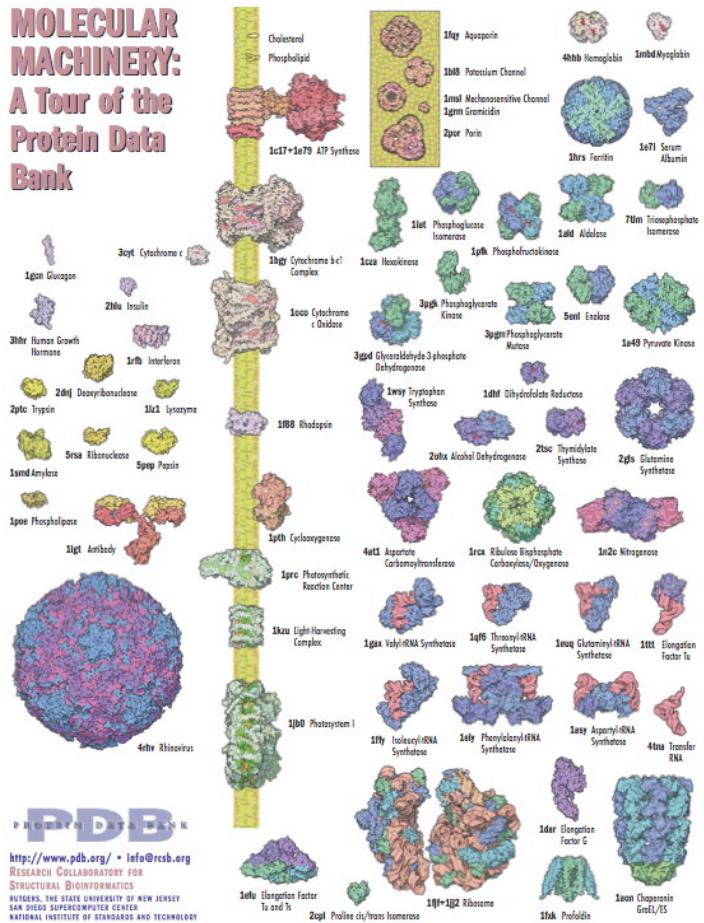


# Concept review: Chromatography (applied to protein purification)

# Proteins come in all shapes and sizes

2

## MOLECULAR MACHINERY: A Tour of the Protein Data Bank



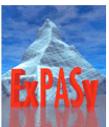
While DNA only really varies in length, proteins can vary in:

- Charge
- Solubility
- Size
- Shape
- Function

- A protein must be purified before its structure and the mechanism of its action can be studied, or before it can be used as an analytical tool. However, because proteins vary in size, charge, and structure, (compare to DNA which is more-or-less always the same) no single method can be applied to the isolation of all proteins.
- To isolate one particular protein from the estimated 10,000 different proteins in a cell is a daunting task that requires methods both for separating proteins and for detecting the presence of specific proteins.
- Any molecule, whether protein, carbohydrate, or nucleic acid, can be separated from other molecules based on differences in some physical characteristic. Although the sequence of amino acids in a protein uniquely determines its function, the most useful physical characteristics for separation of proteins are its charge and size, defined as either length or mass.
- Many thousands of proteins have been purified in active form on the basis of such characteristics as solubility, size, charge, and specific binding affinity. Usually, protein mixtures are subjected to a series of separations, each based on a different property to yield a pure protein. Substantial quantities of purified proteins, of the order of many milligrams, are needed to fully elucidate their three-dimensional structures and their mechanisms of action. Thus, the overall yield is an important feature of a purification scheme. A variety of purification techniques are available.
- In this section, we briefly outline different techniques for separating proteins based on their physical properties. These techniques could also be applied to the separation of other molecules, be they biomolecules or otherwise.

Site Map | Search ExPASy | Contact us

Search  for

 **ExPASy Proteomics Server**

The ExPASy (Expert Protein Analysis System) [proteomics](#) server of the [Swiss Institute of Bioinformatics](#) (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE ([Disclaimer](#) / [References](#)).

[\[Announcements\]](#) [\[Job opening\]](#) [\[Mirror Sites\]](#)

Databases	Tools and software packages
<ul style="list-style-type: none"> <li><a href="#">Swiss-Prot and TrEMBL</a> - Protein knowledgebase</li> <li><a href="#">PROSITE</a> - Protein families and domains</li> <li><a href="#">SWISS-2DPAGE</a> - Two-dimensional polyacrylamide gel electrophoresis</li> <li><a href="#">ENZYME</a> - Enzyme nomenclature</li> <li><a href="#">SWISS-3DIMAGE</a> - 3D images of proteins and other biological macromolecules</li> <li><a href="#">SWISS-MODEL Repository</a> - Automatically generated protein models</li> <li><a href="#">GermOnLine</a> - Knowledgebase on germ cell differentiation</li> <li><a href="#">Ashbya Genome Database</a></li> <li><a href="#">Links to many other molecular biology databases</a></li> </ul>	<ul style="list-style-type: none"> <li><a href="#">Proteomics and sequence analysis tools</a> <ul style="list-style-type: none"> <li><a href="#">Proteomics</a> [<a href="#">Alidente</a> (PMF), <a href="#">Phenyx</a> (MS/MS), <a href="#">FindMod</a>, <a href="#">PeptideMass</a>, ...]</li> <li><a href="#">DNA &gt; Protein</a> [<a href="#">Translate</a>]</li> <li><a href="#">Similarity searches</a> [<a href="#">BLAST</a>]</li> <li><a href="#">Pattern and profile searches</a> [<a href="#">ScanProsite</a>]</li> <li><a href="#">Post-translational modification and topology prediction</a></li> <li><a href="#">Primary structure analysis</a> [<a href="#">ProParam</a>, <a href="#">pI/MW</a>, <a href="#">ProScale</a>]</li> <li><a href="#">Secondary and tertiary structure prediction</a> [<a href="#">SWISS-MODEL</a>, <a href="#">Swiss-PdbViewer</a>]</li> <li><a href="#">Alignment</a> [<a href="#">COFFEE</a>, <a href="#">MSA</a>]</li> <li><a href="#">Biological text analysis</a></li> </ul> </li> <li><a href="#">ImageMaster / Melanie</a> - Software for 2-D PAGE analysis</li> <li><a href="#">MSight</a> - Mass Spectrometry Imager</li> <li><a href="#">Roche Applied Science's Biochemical Pathways</a></li> </ul>
Education and services	Documentation
<ul style="list-style-type: none"> <li><a href="#">The ExPASy FTP server</a></li> <li><a href="#">Swiss-Shop</a> - automatically obtain (by email) new sequence entries relevant to your field(s) of interest</li> <li><a href="#">Vital-IT</a> - The HPC Center for Life Sciences</li> <li><a href="#">e-Proxemis</a>  - Proteomics-oriented Bioinformatics Training Portal</li> <li><a href="#">Master's degree in Proteomics and Bioinformatics</a></li> <li><a href="#">Proteomics courses</a> - two courses covering Separation Sciences &amp; Mass Spectrometry for Proteomics</li> <li><a href="#">Proteomics Core Facility</a> (previously SWISS-2DSERVICE) - get your 2-D Gels performed according to Swiss standards</li> </ul>	<ul style="list-style-type: none"> <li><a href="#">What's New on ExPASy</a></li> <li><a href="#">SWISS-FLASH</a> electronic bulletins</li> <li><a href="#">Swiss-Prot documents</a></li> <li><a href="#">How to create HTML links to ExPASy</a></li> <li><a href="#">Complete table of available documents</a></li> </ul>

- Before you start purifying a protein, it would be useful to have some idea of its physical properties. Generally speaking, if you are interested in purifying a protein then you already know its primary structure (the sequence of amino acids).
- From the sequence alone, you can calculate some properties which can help guide the purification procedure. However, these tools do not account for the fact that the protein has a 3-dimensional structure and only calculates properties that are independent of structure.

Primary structure analysis
<ul style="list-style-type: none"><li><a href="#">ProtParam</a> - Physico-chemical parameters of a protein sequence (amino-acid and atomic compositions, <i>pI</i>, extinction coefficient, etc.)</li><li><a href="#">Compute pI/Mw</a> - Compute the theoretical <i>pI</i> and <i>Mw</i> from a UniProt Knowledgebase entry or for a user sequence</li><li><a href="#">ScanSite pI/Mw</a> - Compute the theoretical <i>pI</i> and <i>Mw</i>, and multiple phosphorylation states</li><li><a href="#">MW, pI, Titration curve</a> - Computes <i>pI</i>, composition and allows to see a titration curve</li><li><a href="#">Radar</a> - De novo repeat detection in protein sequences</li><li><a href="#">REP</a> - Searches a protein sequence for repeats</li><li><a href="#">REPRO</a> - De novo repeat detection in protein sequences</li><li><a href="#">TRUST</a> - De novo repeat detection in protein sequences</li><li><a href="#">SAPS</a> - Statistical analysis of protein sequences at EMBnet-CH [Also available at <a href="#">EBI</a>]</li><li><a href="#">Coils</a> - Prediction of coiled coil regions in proteins (Lupas's method) at EMBnet-CH [Also available at <a href="#">PBIL</a>]</li><li><a href="#">Paircoil</a> - Prediction of coiled coil regions in proteins (Berger's method)</li><li><a href="#">Multicoil</a> - Prediction of two- and three-stranded coiled coils</li><li><a href="#">2ZIP</a> - Prediction of Leucine Zippers</li><li><a href="#">PESTfind</a> - Identification of PEST regions at EMBnet Austria</li><li><a href="#">HLA_Bind</a> - Prediction of MHC type I (HLA) peptide binding</li><li><a href="#">PEPVAC</a> - Prediction of supertypic MHC binders</li><li><a href="#">RANKPEP</a> - Prediction of peptide MHC binding</li><li><a href="#">SYFPEITHI</a> - Prediction of MHC type I and II peptide binding</li><li><a href="#">ProtScale</a> - Amino acid scale representation (Hydrophobicity, other conformational parameters, etc.)</li><li><a href="#">Drawhca</a> - Draw an HCA (Hydrophobic Cluster Analysis) plot of a protein sequence</li><li><a href="#">Protein Colourer</a> - Tool for coloring your amino acid sequence</li><li><a href="#">Three To One</a> - Tool to convert a three-letter coded amino acid sequence to single letter code</li><li><a href="#">Colorseq</a> - Tool to highlight (in red) a selected set of residues in a protein sequence</li><li><a href="#">HelixWheel</a> / <a href="#">HelixDraw</a> - Representations of a protein fragment as a helical wheel</li><li><a href="#">RandSeq</a> - Random protein sequence generator</li></ul>

- As can be seen in this table, the types of parameters calculated by primary analysis tools include:
  - molecular weight
  - amino acid/atomic composition
  - $pI$  = isoelectric point, or the pH at which the number of positive charges = number of negative charges. There are necessarily some assumptions here, since the 3D structure could influence the  $pKa$  of individual ionizable groups. The calculated value should be reliable for the denatured protein.
  - extinction coefficient at 280 nm. This calculation is based on the number of Trp + Tyr in the protein. This calculation assumes that the protein is denatured so that 3D structure is not influencing the properties of these chromophores.
  - other tools look for patterns in the amino acids that may indicate repetitive common structures (secondary structure prediction)
  - hydrophobicity as a function of residue number. This might be useful for indicating the presence of transmembrane helices, for example.

# ProtParam

User-provided sequence:

```
1 MVSKGEEELFV 11 GVVPILVLED 21 GDVNCHKFVSV 31 SGEGEQDATY 41 GKLTLLKFICT 51 TGKLPVWPWT
61 LVTTTLTYGVQ 111 CFSRYPDHMK 211 QHDFFKSAMF 311 EGVVQERTIF 411 FKDDGNYKTR 511 AEVKFEGDTL
121 VNRIELKGID 1211 FKEDGQNLCH 221 KLEYNNYNSHN 321 VYIMADKQKN 421 GIKVNFKIRH 521 NIEDGSVOLA
181 DHYQONTPIG 1811 DGPVLLPDNH 281 YLSTQASLSK 381 DPNEKRDHMV 481 LLEFVTAAGI 581 TLGMDELYK
```

[References](#) and [documentation](#) are available.

Number of amino acids: 239

Molecular weight: 26941.4

Theoretical pI: 5.58

Amino acid composition:

Ala (A)	8	3.3%
Arg (R)	6	2.5%
Asn (N)	13	5.4%
Asp (D)	18	7.5%
Cys (C)	2	0.8%
Gln (Q)	8	3.3%
Glu (E)	16	6.7%
Gly (G)	22	9.2%
His (H)	9	3.8%
Ile (I)	12	5.0%
Leu (L)	21	8.8%
Lys (K)	20	8.4%
Met (M)	6	2.5%
Phe (F)	12	5.0%
Pro (P)	10	4.2%
Ser (S)	10	4.2%
Thr (T)	16	6.7%
Trp (W)	1	0.4%
Tyr (Y)	11	4.6%
Val (V)	18	7.5%

Total number of negatively charged residues (Asp + Glu): 34

Total number of positively charged residues (Arg + Lys): 26

Atomic composition:

Carbon	C	1208
Hydrogen	H	1865
Nitrogen	N	317
Oxygen	O	366
Sulfur	S	8

Formula: C<sub>1208</sub>H<sub>1865</sub>N<sub>317</sub>O<sub>366</sub>S<sub>8</sub>  
Total number of atoms: 3764

<http://ca.expasy.org/tools/protparam.html>

5

## Extinction coefficients:

Conditions: 6.0 M guanidium hydrochloride  
0.02 M phosphate buffer  
pH 6.5

Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>.

The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276	278	279	280	282
Ext. coefficient	21495	21127	20575	19890	18920
Abs 0.1% (=1 g/l)	0.798	0.784	0.764	0.738	0.702

	276	278	279	280	282
Ext. coefficient	21350	21000	20455	19770	18800
Abs 0.1% (=1 g/l)	0.792	0.779	0.759	0.734	0.698

## Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).  
>20 hours (yeast, in vivo).  
>10 hours (Escherichia coli, in vivo).

## Instability index:

The instability index (II) is computed to be 29.06  
This classifies the protein as stable.

Aliphatic index: 79.04

Grand average of hydropathicity (GRAVY): -0.467

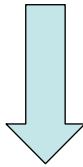
- Here is the output from a Protparam analysis of the GFP protein sequence provided earlier.
- I think that the output is largely self-explanatory, though you should also read the documentation to understand the parameters.
- <http://ca.expasy.org/tools/protparam-doc.html>

# Initial steps of a routine protein purification

6

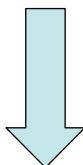
## Source of protein

- Tissue
- Bacteria (endogenous)
- Bacteria (recombinant)



## Homogenization/lysis

- Sonification
- French press
- Detergents
- And many more...



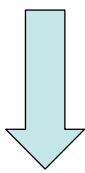
## French Pressure cell vs. French Press

Nice overview from Pierce at 'Protein Methods Library' (long URL, Google it)

- There are many different disruption methods, some mechanical and some chemical. Proteases will be liberated upon cell disruption, thus it is necessary to take precautions to prevent digestion of your target protein.
  - 1. Cell disruption should be performed at cold temperatures. Keep the sample on ice as much as possible and use chilled solutions. This will decrease the activity of the proteases for the simple reasons that all chemical reactions occur more slowly at low temperature.
  - 2. Add protease inhibitors to specifically inactivate the proteases. Cocktails of inhibitors are available from a variety of suppliers.
  - 3. Work quickly. Try to minimize the time that your protein of interest is in the presence of the proteases.
  - 4. Alternatively, the source material can be disrupted under denaturing conditions such as 8 M Urea or guanidinium HCl. These conditions will cause all proteins to unfold, including both the proteases and your protein of interest. This means that no proteolysis will occur but you will also have to refold your protein at some later time. Sometimes this is easy, sometimes not so easy...
- The most common method for disrupting (lysis) of bacterial cells is the use of the French Press (not the coffee maker!). The design can be a bit misleading, since it generally suggests to the novice that the cells are disrupted through the subjecting them to high pressure. This is only part of the story. The cells are actually forced past a small nylon bead at the exit of the inner chamber. The rapid change in pressure is what causes the cells to pop.
- Following lysis, it is necessary to separate soluble proteins from insoluble cellular material by centrifugation. The centrifugal force and duration of centrifugation can be adjusted to ensure that the insoluble materials sediment into a pellet. If your protein of interest is a soluble protein, you would essentially want to pellet all of the insoluble material so it is a good idea to spin as fast as your centrifuge safely allows you.
- In cases where the goal is to enrich a eukaryotic cell lysate for certain organelles, it is possible to spin at speeds where only the largest organelles sediment. At the lowest speeds, only nuclei will sediment. At higher speeds it will be mitochondria. And at yet higher speeds it will be fragments of the endoplasmic reticulum known as microsomes. The supernatant fraction still contains a large mixture of proteins, which can be collected by decanting the supernatant and then subjecting it to further purification methods.

# Salting out and dialysis

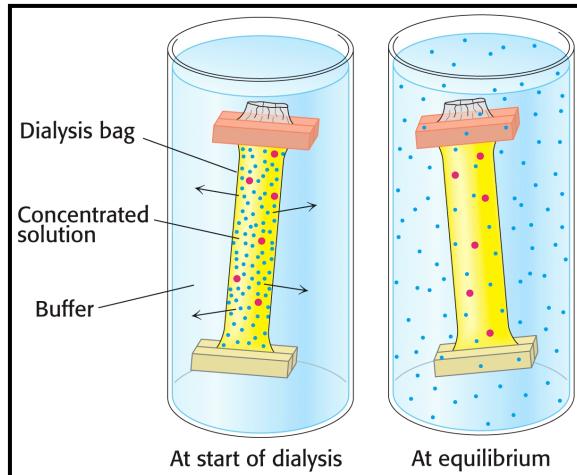
7



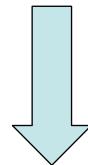
$(\text{NH}_4)_2\text{SO}_4$   
precipitation for  
crude  
fractionation of  
proteins.



Dialysis to separate protein from small molecules such as salt and metabolites



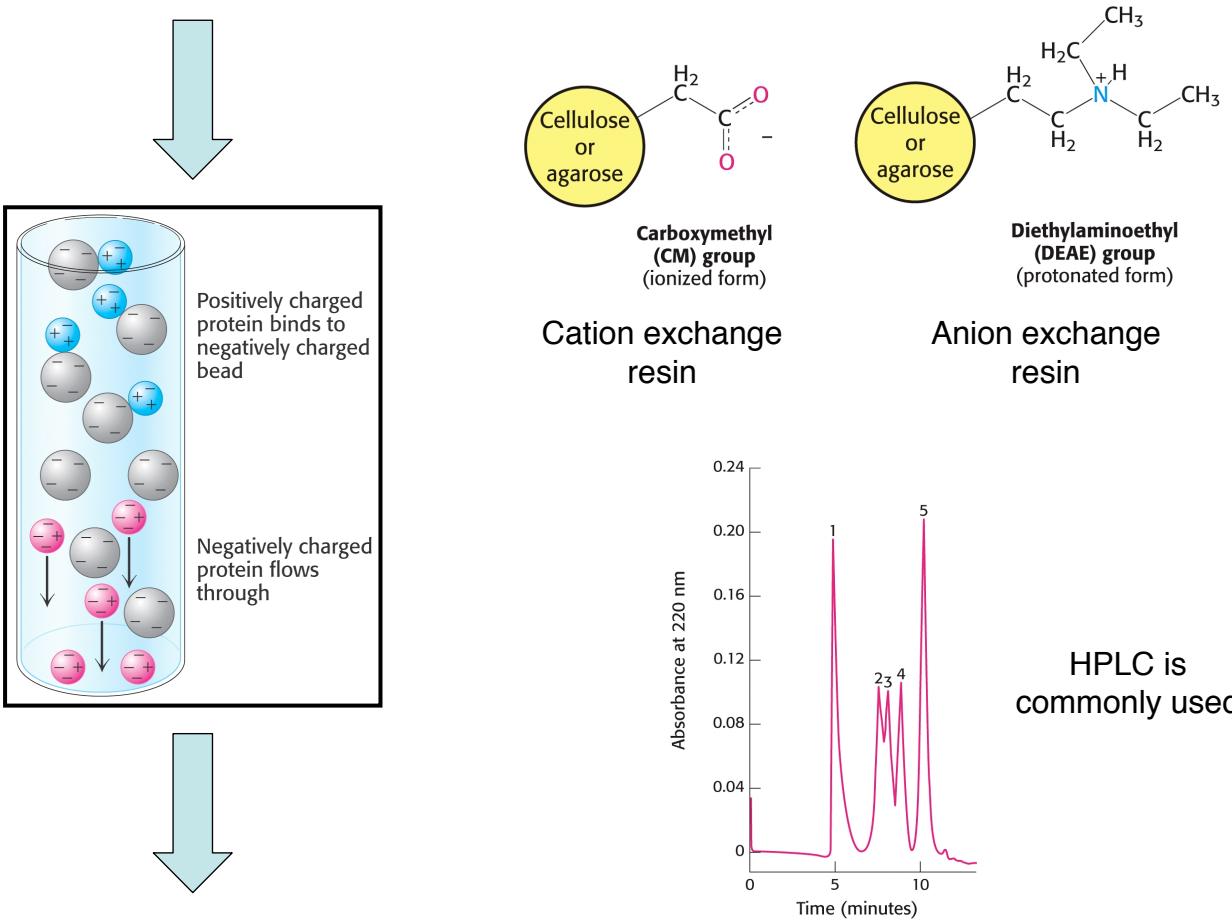
- You will need to know which fraction contains the protein of interest. This requires an assay, a topic which will be covered later in the course.



- **Salting Out.** Most proteins are less soluble at high salt concentrations, an effect called salting out. The salt concentration at which a protein precipitates differs from one protein to another. Hence, salting out can be used to fractionate proteins. For example, 0.8 M ammonium sulfate precipitates fibrinogen, a blood-clotting protein, whereas a concentration of 2.4 M is needed to precipitate serum albumin. Salting out is also useful for concentrating dilute solutions of proteins. Dialysis can be used to remove the salt if necessary. This is not a technique that is routinely done in the age of molecular biology because it generally provides only a slight (2-10-fold) enrichment of the protein of interest. Could be useful in certain cases though...
- **Dialysis.** Proteins can be separated from small molecules by dialysis through a semipermeable membrane, such as a cellulose membrane with pores. Molecules having dimensions significantly greater than the pore diameter are retained inside the dialysis bag, whereas smaller molecules and ions traverse the pores of such a membrane and emerge in the dialysate outside the bag. This technique is useful for removing a salt or other small molecule, but it will not distinguish between proteins effectively. Dialysis tubing is characterized by the molecular weight cut off (MWCO). The MWCO describes the molecular weight at which a compound will be 90% retained following overnight (17-hour) dialysis. The MWCO is determined by testing many different proteins of known molecular weight. In general, these MWCO apply to globular molecules, such as most proteins. More linear proteins may be able to pass through the pores, even though their molecular weight exceeds the stated MWCO. To compensate for this, choose a dialysis device with a smaller MWCO. For DNA or RNA, a MWCO no greater than one-third the MW should be used in order to prevent excessive sample loss. [From Pierce technical resource]. Note that the final concentration of freely diffusing molecules will be the same inside and outside the bag. The larger the volume of the dialysis buffer, the lower the final concentration of the freely diffusing molecules.

# Ion-exchange chromatography

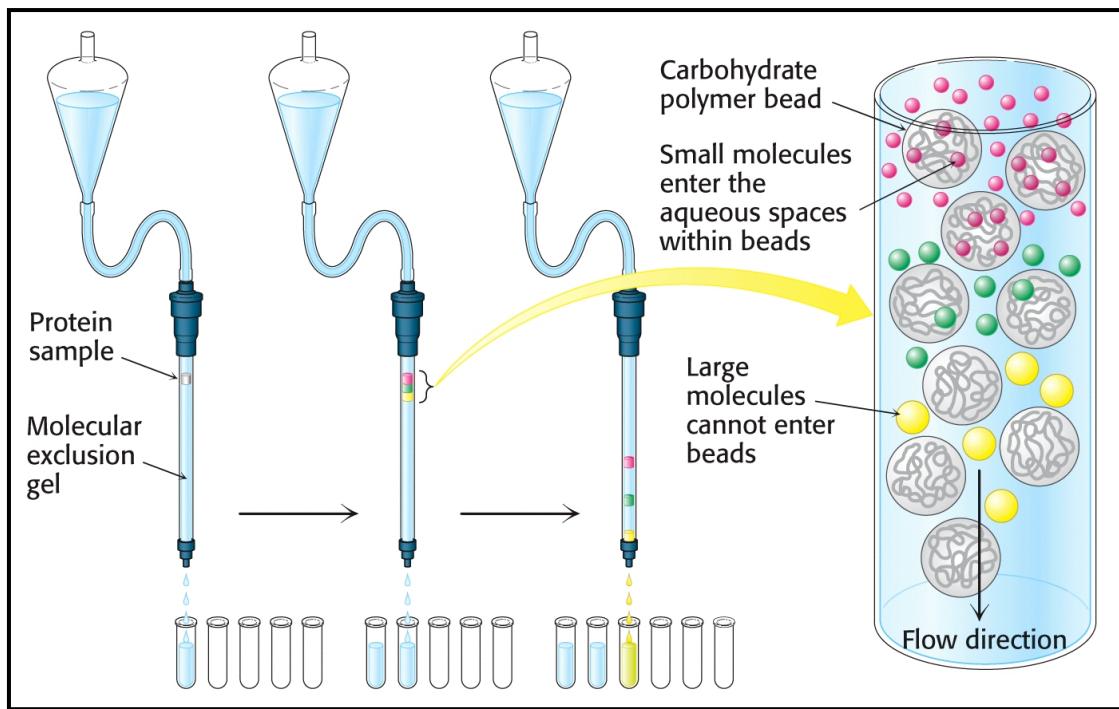
8



- **Ion-Exchange Chromatography.** Proteins can be separated on the basis of their net charge by ion-exchange chromatography. If a protein has a net positive charge at pH 7 (i.e. a pI of >7), it will usually bind to a column of beads containing carboxylate groups, whereas a negatively charged protein will not. Whether a protein has a positive, negative, or neutral charge depends on both the pI of the protein (an intrinsic property of the protein) and the pH of the solution. It is generally preferable to keep the pH within a pH unit or 2 of physiological pH (7.4). These considerations will dictate whether cation or anion exchange is used and the pH of the running buffer.
- A positively charged protein bound to such a column can then be eluted (released) by increasing the concentration of sodium chloride or another salt in the eluting buffer because sodium ions compete with positively charged groups on the protein for binding to the column. Alternatively, a change in the pH could be used in some cases.
- Proteins that have a low density of net positive charge will tend to emerge first, followed by those having a higher charge density. Positively charged proteins (cationic proteins) can be separated on negatively charged carboxymethyl-cellulose (CM-cellulose) columns (cation exchange resins). Conversely, negatively charged proteins (anionic proteins) can be separated by chromatography on positively charged diethylaminoethyl-cellulose (DEAE-cellulose) columns (anion exchange resins).
- The isoelectric point (pI) of a protein is the pH at which its net charge is zero. In very general terms a protein will stick to a cation exchange resin below its pI and to an anion exchange resin above its pI. However, what really matters is the local charge distribution on the surface of the protein (i.e. patches of residues with similar charges). One might gain some insight into the charge distribution by looking at the 3D structure.

# Gel-filtration chromatography

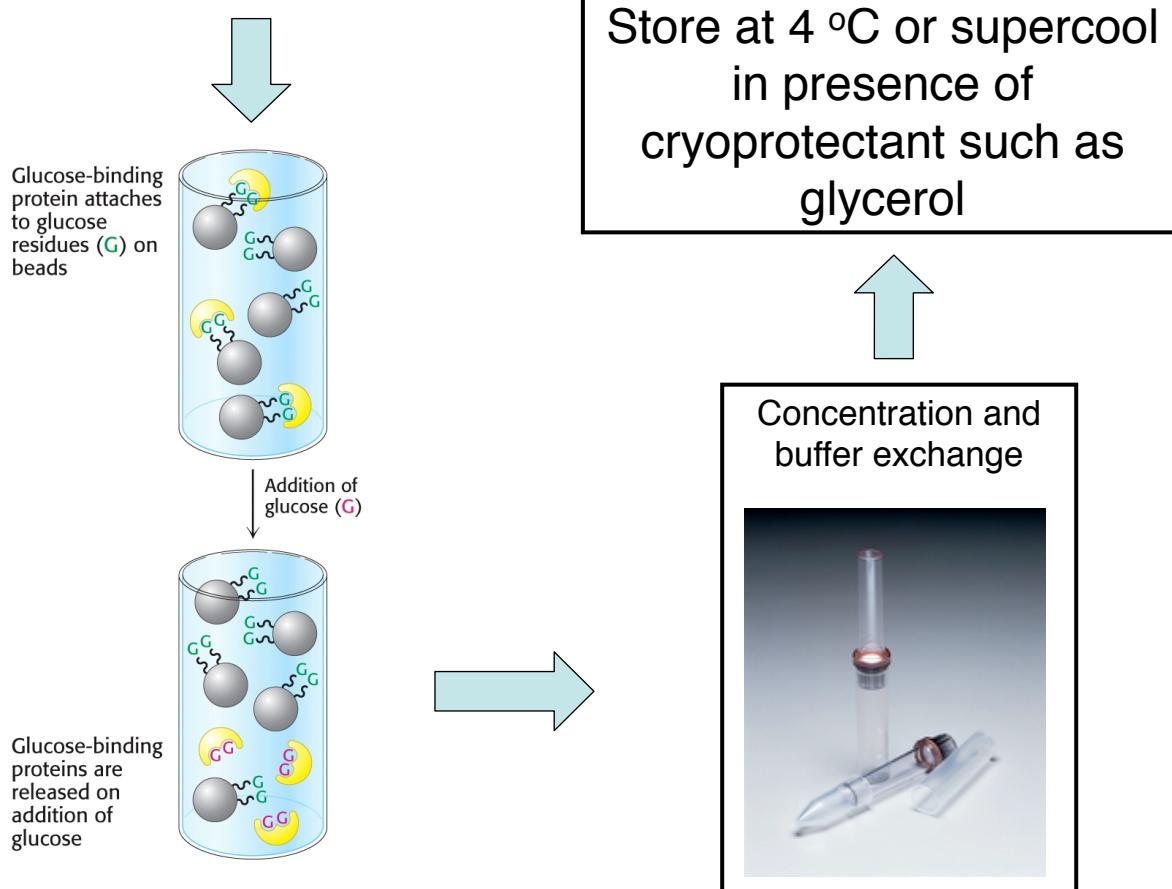
9



- **Gel-Filtration Chromatography.** Separations on the basis of size can be achieved by the technique of gel-filtration chromatography. The sample is applied to the top of a column consisting of porous beads made of an insoluble but highly hydrated polymer such as dextran or agarose (which are carbohydrates) or polyacrylamide.
- Sephadex, Sepharose, and Bio-gel are commonly used commercial preparations of these beads, which are typically 100  $\mu\text{m}$  (0.1 mm) in diameter. Small molecules can enter these beads, but large ones cannot. The result is that small molecules are distributed in the aqueous solution both inside the beads and between them, whereas large molecules are located only in the solution between the beads.
- Large molecules flow more rapidly through this column and emerge first because a smaller volume is accessible to them. Molecules that are of a size to occasionally enter a bead will flow from the column at an intermediate position, and small molecules, which take a longer, tortuous path, will exit last.
- *Q: Can you explain a little more about gel filtration chromatography?*
- *A: Gel-filtration chromatography is a separation technique that is based on differences in molecular size. The separation media is porous beads that contain pores of many different sizes. Molecules passing through the separation media can enter the pores on the beads. The key to the separation is that smaller molecules can enter more of the pores than the larger molecules. This means that the smaller the molecule, the more column volume it has to pass through. Another way of thinking of this is that the smaller molecules have to follow a longer path to reach the end of the column. Accordingly, larger molecules exit the column before smaller molecules. This technique is widely used for separating proteins, though it is not a high resolution technique and only really effective at separating proteins with large differences in molecular weight.*

# Affinity chromatography and storage

10



- **Affinity Chromatography.** Affinity chromatography is another powerful and generally applicable means of purifying proteins. This technique takes advantage of the high affinity of many proteins for specific chemical groups. For example, the plant protein concanavalin A can be purified by passing a crude extract through a column of beads containing covalently attached glucose residues. Concanavalin A binds to such a column because it has affinity for glucose, whereas most other proteins do not. The bound concanavalin A can then be released from the column by adding a concentrated solution of glucose. The glucose in solution displaces the column-attached glucose residues from binding sites on concanavalin A.
- In general, affinity chromatography can be effectively used to isolate a protein that recognizes group X by (1) covalently attaching X or a derivative of it to a column, (2) adding a mixture of proteins to this column, which is then washed with buffer to remove unbound proteins, and (3) eluting the desired protein by adding a high concentration of a soluble form of X or altering the conditions to decrease binding affinity. Affinity chromatography is most effective when the interaction of the protein and the molecule that is used as the bait is highly specific.
- This is often used as a last step simply because the columns are more expensive or precious because they are custom made. Protein that has been through ion-exchange and gel filtration chromatography is relatively 'clean' and thus will probably not contaminate the affinity column. Of course, if the affinity column is not particularly expensive and not precious, you might as well use it as a first step in the purification.
- As we will see later, in the modern age of molecular biology, most protein purifications are based almost solely on some type of affinity chromatography.

# 5 'golden rules' of working with proteins

11

**Golden rule #1:** Keep it buffered! Solutions should always be buffered somewhere between 6.5-7.5.

**Golden rule #2:** No bubbles! Do not agitate a protein solution!  
Think of milk - proteins like to form bubbles!

**Golden rule #3:** Keep it on ice! Work at 4 °C!

**Golden rule #4:** Keep it reduced! Always maintain 1-2 mM reducing agent in your buffer.

**Golden rule #5:** Freeze fast, thaw fast! 10%+ glycerol is necessary when freezing (supercooling) proteins in liq. N<sub>2</sub> or dry ice/acetone. Running luke warm tap water is best for a fast thaw.

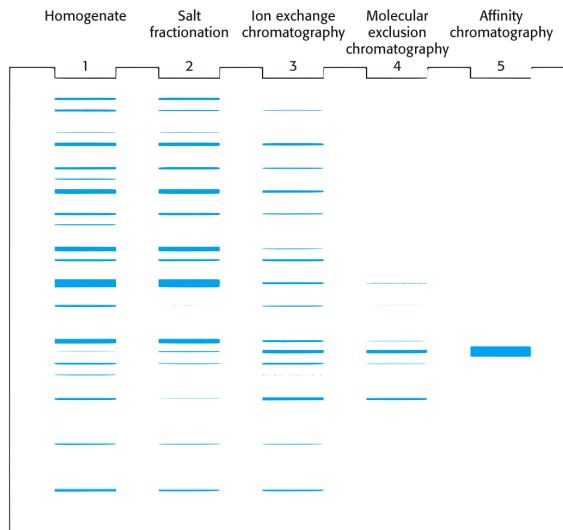
- Your enemies are pH extremes, bubbles, proteases, oxygen, and ice crystals. Avoid them all!

# Assessing a protein purification protocol

12

**TABLE 4.1** Quantification of a purification protocol for a fictitious protein

Step	Total protein (mg)	Total activity (units)	Specific activity, (units mg <sup>-1</sup> )	Yield (%)	Purification level
Homogenization	15,000	150,000	10	100	1
Salt fractionation	4,600	138,000	30	92	3
Ion-exchange chromatography	1,278	115,500	90	77	9
Molecular exclusion chromatography	68.8	75,000	1,100	50	110
Affinity chromatography	1.75	52,500	30,000	35	3,000



## To fully quantify a protein purification you require:

- Total protein
- Total activity
- Specific activity (activity/mg protein)
- SDS-PAGE gel (discussed later)

- To determine the success of a protein purification scheme, we monitor the procedure at each step by determining specific activity and by performing an SDS-PAGE analysis.
- Total protein. The quantity of protein present in a fraction is obtained by determining the protein concentration of a part of each fraction and multiplying by the fraction's total volume.
- Total activity. The enzyme activity for the fraction is obtained by measuring the enzyme activity in the volume of fraction used in the assay and multiplying by the fraction's total volume.
- Specific activity. This parameter is obtained by dividing total activity by total protein.
- Yield. This parameter is a measure of the activity retained after each purification step as a percentage of the activity in the crude extract. The amount of activity in the initial extract is taken to be 100%.
- Purification level. This parameter is a measure of the increase in purity and is obtained by dividing the specific activity, calculated after each purification step, by the specific activity of the initial extract.
- The SDS-PAGE in shows that, if we load a constant amount of protein onto each lane after each step, the number of bands decreases in proportion to the level of purification, and the amount of protein of interest increases as a proportion of the total protein present.